



# Traitement du signal pour les prothèses visuelles: approche biométrique et sensori-motrice

Barthélémy Durette

## ► To cite this version:

Barthélémy Durette. Traitement du signal pour les prothèses visuelles: approche biométrique et sensori-motrice. Traitement du signal et de l'image [eess.SP]. Université Joseph-Fourier - Grenoble I, 2009. Français. NNT: . tel-00537921

**HAL Id: tel-00537921**

**<https://theses.hal.science/tel-00537921>**

Submitted on 19 Nov 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Université Grenoble 1 - Joseph Fourier

N° attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

## **THESE**

*pour obtenir le grade de*

**Docteur de l'Université Grenoble I**

*Spécialité :*

Ingénierie de la Cognition, de l'Interaction, de l'Apprentissage et de la Création  
option : Sciences Cognitives

dans le cadre de l'Ecole Doctorale :

Ingénierie pour la Santé, la Cognition et l'Environnement

*présentée et soutenue publiquement par*

**Barthélémy Durette**

*le 10 Juillet 2009*

Titre :

**TRAITEMENT DU SIGNAL POUR LES PROTHÈSES VISUELLES:  
APPROCHE BIOMIMÉTIQUE ET SENSORI-MOTRICE.**

Jury :

Edwige Pissaloux  
Charles Lenay  
Philippe Gaussier  
Kevin O'Regan

Présidente du Jury  
Rapporteur  
Rapporteur  
Examineur

Professeur, ISIR/Université Pierre et Marie Curie  
Professeur, Université Technologique de Compiègne  
Professeur, ENSEA  
DR CNRS, Institut Paris Descartes, Neurosciences de la  
Cognition  
Maître de Conférences, Université Paris Descartes  
Professeur émérite, Université Joseph Fourier, Grenoble  
CR CNRS, Université Pierre Mendès-France, Grenoble

Sylvain Hanneton  
Jeanny Hérault  
David Alleysson

Examineur  
Directeur de thèse  
Codirecteur de thèse



# Résumé

Bien qu'elles existent depuis plus de quarante ans, les prothèses visuelles, qu'elles soient invasives (implants) ou non invasives (systèmes de substitution sensorielle), n'ont pas percé dans le secteur du handicap. Il serait difficile d'imputer cet état de fait à des limitations technologiques : depuis les premières approches, les prothèses visuelles n'ont cessé de se perfectionner et de se diversifier. Toutefois, si la question de savoir *comment* transmettre le signal est bien documentée, la question de savoir *quel signal transmettre* a été plus rarement abordée.

Dans cette thèse, nous identifions des outils susceptibles d'étendre la conception des dispositifs de suppléance visuelle au-delà de la seule question de l'interface. Nous puisons pour cela dans les développements récents sur les théories actives de la vision et dans les implications de ces théories sur la conception et l'évaluation de dispositifs de perception. Nous nous inspirons par ailleurs du fonctionnement des systèmes perceptifs naturels, et plus particulièrement du traitement de l'information dans le système rétino-cortical humain, dont nous proposons une simulation, implémentée dans le dispositif de substitution sensorielle visuo-auditif TheVIBE.

L'impact de ce traitement biomimétique du signal est testé expérimentalement sur des sujets humains équipés du dispositif dans différentes tâches de localisation et de mobilité. L'observation du comportement des sujets témoigne de l'intérêt du traitement du signal dans les prothèses visuelles et apporte un nouveau regard sur son rôle potentiel dans les dispositifs de perception, qu'ils soient artificiels ou naturels.

# Summary

Although the development of invasive (implants) or non-invasive (sensory substitution system) visual prostheses started 40 years ago, the use of such devices in the visually-impaired community is still anecdotal. Technological means are not likely to account for this situation. Since the former approaches, the means to interface a video-camera with the brain are more and more sophisticated and diversified. The question about *how to transmit the signal* is thus well documented. However, the question about *what signal should be transmitted* is rarely addressed.

In this thesis, we will identify tools to push the conception of visual prosthesis beyond the scope of mere interfacing. Active theories of vision are our guideline to do so. Moreover, our work is inspired by natural visual systems : from a model of the human' early visual system established in our laboratory, we have derived a simulation software and implemented it in a recent visuo-auditory sensory substitution system named TheVIBE.

*Our approach is validated with human subjects wearing the device under different tasks requiring localization and/or mobility. Observation of the subject's behavior when using the device demonstrates the interest of signal processing in a visual prosthesis, and give new trails concerning its role in the visual process, be it artificial or natural.*



# Introduction

Bien qu'ils existent depuis plus de trente ans, les systèmes de suppléance visuelle par des aides optiques et électroniques n'ont pas percé dans le secteur du handicap (Lenay, 2003 ; Gapenne et Gaussier, 2005). Les résultats obtenus avec les premiers dispositifs étaient pourtant encourageants : immersion rapide, mise en extériorité du percept, capacités de discrimination sommaires (Bach-y-Rita, 1969). Sur le plan scientifique, ces dispositifs ont permis de mettre en évidence l'importance de l'action dans le processus perceptif et ont soutenu l'émergence concomitante des théories actives de la perception dans le domaine de la psychologie cognitive. *Sur le plan du handicap par contre*, ils n'ont eu qu'un très faible impact : leur utilisation quotidienne par des sujets aveugles est restée très marginale.

Il semble difficile d'imputer cet échec à des limitations technologiques : depuis les premières approches, les dispositifs de suppléance n'ont cessé de se perfectionner et de se diversifier. Dans l'engouement général pour créer des systèmes de plus en plus élaborés, il semble toutefois qu'une question fondamentale ait été laissée de côté : la question de la *pertinence* et de la plausibilité phénoménologique du signal envoyé via ces systèmes. Le signal que l'on envoie est-il vraiment adapté à l'information visuelle et à la manière dont fonctionne le système visuel ? Il s'agit moins de savoir comment transmettre le signal que de *savoir quel signal transmettre* et sous quel format : comment le signal en provenance de la caméra doit-il être codé pour que le sujet puisse extraire efficacement de l'information à l'aide du dispositif ?

L'objectif de cette thèse est d'identifier les outils susceptibles de pousser la conception des dispositifs de suppléance visuelle au delà de la seule question de l'interface. Nous puisons pour cela dans les développements récents sur les théories actives de la vision, explicités notamment dans la théorie des contingences sensori-motrices (O'Regan et Noë, 2001), et dans les implications de ces théories sur la conception et l'évaluation des dispositifs de perception (Lenay, 2003; Auvray, 2004 ; Gapenne et Gaussier, 2005). Nous nous inspirons par ailleurs du fonctionnement des systèmes perceptifs naturels, et plus particulièrement du traitement de l'information dans le système rétinocortical humain, dont nous proposons un modèle, issue de plusieurs travaux de thèse au laboratoire, et une simulation logicielle.

Notre approche est centrée sur le degré de pertinence des signaux transmis par les systèmes de suppléance visuelle. Dans un cadre sensori-moteur en particulier, la qualité fondamentale du signal ne réside pas dans sa fidélité au monde mesuré, mais dans la *richesse du couplage* qu'il introduit entre l'homme et son environnement visuel. Il s'agit donc d'enrichir ce couplage. Notre hypothèse est que c'est précisément le rôle que jouent les circuits de la rétine et les projections corticales chez l'Homme, bien au-delà du simple processus de capture et de transport de l'information lumineuse jusqu'au cortex.

Ce document est organisé en cinq chapitres :

## **Chapitre 1 : Voir et restaurer la vision : du traitement de l'information au couplage sensori-moteur**

Ce chapitre posera les bases de notre travail, aussi bien sur le plan pratique (état de l'art sur les prothèses visuelles et le traitement du signal dans ces prothèses), que sur le plan théorique (approche sensori-motrice de la perception et conséquences sur la conception d'une prothèse)

## **Chapitre 2 : Simulation de la chaîne de traitement rétino-corticale du primate**

Nous présenterons ici notre source d'inspiration : les premières aires de traitement du système visuel du primate. Après une description succincte de la physiologie de la chaîne rétino-corticale, nous présenterons le modèle développé au laboratoire ainsi que le logiciel de simulation qui a été réalisé dans le cadre de cette thèse.

## **Chapitre 3 : Configuration du dispositif TheVIBE**

Nous aborderons ici les questions concernant l'échantillonnage de l'image dans le dispositif de substitution que nous utilisons, ainsi que les choix réalisés en terme de codage sonore. Nous verrons en particulier dans ces deux cas l'influence que peut jouer l'approche sensorimotrice sur ces éléments "de base" de la conception d'un dispositif.

## **Chapitre 4 : Premier pas avec le dispositif**

Nous verrons ici les premiers tests comportementaux réalisés avec le dispositif. Pour démarrer, nous avons choisi des tâches élémentaires :

1. Diriger la caméra du dispositif vers une cible lumineuse
2. Aller chercher une cible à différents endroits dans une pièce

Dans les deux cas, nous avons étudié l'influence d'un échantillonnage spatialement variant biomimétique de l'image sur les performances et sur la stratégie des sujets.

## **Chapitre 5 : Evaluation de l'efficacité du dispositif de substitution pour la mobilité**

Dans ce chapitre, nous aborderons la question de l'intérêt du traitement spatio-temporel opéré au niveau de la rétine. On identifie en particulier plusieurs voies à la sortie de la rétine. La voie Magnocellulaire est réputée transmettre l'information liée au contexte et au mouvement. Nous avons donc choisi de tester la simulation de cette voie dans le dispositif dans une tâche de mobilité.

Nous dresserons enfin un résumé des contributions de cette thèse et proposerons un certain nombre de perspectives, tant sur le plan appliqué que sur le plan de la recherche fondamentale.

# Table des matières

Chapitre 1 : Voir et restaurer la vision, du traitement de l'information au couplage sensori-moteur.....	7
1.1 Les prothèses visuelles .....	7
1.1.1 Les implants.....	8
1.1.2 La substitution sensorielle.....	11
1.2 Le traitement du signal dans les prothèses visuelles.....	13
1.3 Qu'est-ce que voir : une question très actuelle.....	15
1.3.1 Qu'est-ce que voir : les approches actives de la perception visuelle.....	16
1.3.2 La théorie des contingences sensori-motrices.....	17
1.4 Approche sensori-motrice et prothèses visuelles.....	18
1.4.1 Ergonomie sensori-motrice d'une prothèse visuelle.....	18
1.4.2 Immersion dans un dispositif de substitution : un cadre pour l'évaluation.....	21
1.4.3 Une méthodologie d'étude des processus perceptifs.....	21
Références.....	23
Chapitre 2 : Simulation de la chaîne de traitement rétino-corticale du primate.....	29
2.1 Anatomie générale de la chaîne rétino-corticale du primate.....	29
2.1.1 L'œil.....	31
2.1.2 La rétine.....	32
2.1.3 Le corps genouillé latéral.....	36
2.1.4 Les radiations optiques et la projection sur le cortex visuel primaire.....	37
2.2 Comportement spatio-temporel de la rétine, genèse des voies "Magno" et "Parvo".....	38
2.2.1 Modèle générique d'une couche de cellules rétinienne.....	39
2.2.2 Modélisation de la couche plexiforme externe.....	40
2.2.3 Modélisation de la couche plexiforme interne.....	42
2.2.4 Rôle du comportement spatio-temporel de la rétine.....	46
2.3 Dynamique des cellules de la rétine et adaptations locales.....	46
2.3.1 Dynamique du photorécepteur et égalisation locale de la luminance.....	47
2.3.2 Dynamique des cellules ganglionnaires et égalisation locale du contraste.....	48
2.3.3 Rôle de la dynamique des cellules rétinienne.....	48
2.4 Echantillonnage spatialement variant et projections corticales.....	49
2.4.1 Modélisation de la magnification corticale.....	50
2.4.2 Projection polaire sur le cortex.....	52
2.4.3 Rôle de la projection rétino-corticale.....	52
2.5 Réalisation d'un logiciel de simulation de la chaîne rétino-corticale du primate .....	54
Références.....	55



Chapitre 3 : Configuration du dispositif TheVIBE.....	59
3.1 Les dispositifs de substitution visuo-auditive.....	59
3.2 Le dispositif TheVIBE.....	60
3.2.1 Principe de la conversion visuo-sonore.....	60
3.2.2 Configuration de la conversion visuo-sonore.....	62
3.2.3 Les autres paramètres du logiciel.....	63
3.3 Echantillonnage de l'image.....	63
3.3.1 Sous-échantillonnage : ce que dit le traitement du signal.....	64
3.3.2 Création de rétines par échantillonnage aléatoire uniforme.....	67
3.3.3 Création de rétines par division de l'image en rectangles élémentaires.....	69
3.3.4 Création de rétines par auto-organisation de Kohonen.....	70
3.3.5 Echantillonnage spatialement variant.....	78
3.3.6 Correction de l'optique de la caméra.....	79
3.3.7 Conclusion.....	79
3.4 Codage sonore.....	80
3.4.1 Le système auditif humain.....	80
3.4.2 Génération du signal sonore dans TheVIBE.....	82
3.4.3 Choix des paramètres des signaux élémentaires.....	83
3.4.4 Codage de l'intensité sonore.....	84
3.4.5 Codage du panoramique de la source.....	87
3.4.6 Codage de la tonie.....	90
3.4.7 Choix du nombre de champs récepteurs et du taux de rafraichissement de l'image.....	92
3.4.8 Aspects temporels : choix de la fréquence de rafraichissement de la caméra et de la durée des échantillons sonores.....	103
3.4.9 Choix de la fonction de conversion visuo-sonore.....	106
Références.....	111
Chapitre 4 : Premiers pas avec le dispositif TheVIBE.....	115
4.1 Saccades vers une cible à l'aide du dispositif TheVIBE.....	116
4.1.1 Méthode.....	116
4.1.2 Participants et plan d'expérience.....	121
4.1.3 Résultats.....	122
4.1.4 Discussion.....	124
4.1.5 Conclusion.....	128
4.2 Localisation-préhension à l'aide du dispositif TheVIBE.....	129
4.2.1 Méthode.....	129
4.2.2 Protocole.....	132
4.2.3 Participants et plan d'expérience.....	133
4.2.4 Résultats.....	133
4.2.5 Discussion.....	134
4.2.6 Conclusion.....	139
4.3 Conclusion générale.....	139
Références.....	140

Chapitre 5 : Evaluation de l'efficacité du dispositif de substitution pour la mobilité.....	143
5.1 Pourquoi la mobilité?.....	143
5.2 Evaluer la mobilité.....	144
5.2.1 Evaluer un dispositif de substitution : la question de l'apprentissage.....	144
5.2.2 Méthode.....	146
5.3 Expérience 1 : Validation de la méthode et du dispositif TheVIBE.....	149
5.3.1 Participants .....	149
5.3.2 Appareillage.....	149
5.3.3 Protocole.....	150
5.3.4 Résultats.....	151
5.3.5 Discussion.....	155
5.3.6 Conclusion.....	156
5.4 Expérience 2 : effet d'un traitement spatio-temporel de l'image sur l'efficacité du dispositif de substitution.....	158
5.4.1 Participants.....	159
5.4.2 Appareillage.....	159
5.4.3 Protocole.....	161
5.4.4 Résultats.....	161
5.4.5 Discussion.....	164
5.4.6 Conclusion.....	175
Références.....	176
Chapitre 6 : Conclusions et Perspectives.....	178
6.1 Conclusions.....	178
6.1.1 Contributions techniques.....	178
6.1.2 Contributions sur la question du handicap.....	179
6.1.3 Contributions sur le plan de la recherche .....	179
6.2 Perspectives.....	180
6.2.1 Sur le plan du handicap.....	180
6.2.2 Sur le plan de la recherche.....	181
Références.....	182
ANNEXES.....	184
Annexe du chapitre 2.....	186
Annexe du chapitre 3.....	192
Annexe du chapitre 4.....	194
Annexe du chapitre 5.....	204



# Chapitre 1 : Voir et restaurer la vision, du traitement de l'information au couplage sensori-moteur

"Processing and conversion of images for the optimum benefit to the user of a visual prosthesis are areas to which considerable time and effort must be directed in future experiments"

[Dobelle *et al.* 1976]

L'approche que nous avons choisi dans cette thèse pour aborder la question du traitement du signal dans les prothèses visuelles consiste à considérer que, si l'on veut créer un dispositif dont les fonctions sont similaires au système visuel humain, le plus simple est de s'attacher à comprendre en détail comment ce dernier fonctionne, pour ensuite tâcher de reproduire ce mode de fonctionnement.

Cette approche, appelée parfois "reverse engineering" s'est avéré particulièrement pertinente dans le domaine des Sciences Cognitives. C'est sur ce mode de recherche en particulier que Jeanny Hérault a développé au laboratoire un modèle des premières aires de traitement de l'information visuelle. Ce modèle permet, à partir de données physiologiques de base, de simuler les processus de traitement de l'information sur le signal visuel avant son arrivée dans le cortex. L'idée à l'origine de cette thèse était d'implémenter cette simulation dans une prothèse visuelle afin d'en améliorer l'efficacité.

L'intérêt d'une telle idée n'est pas seulement appliquée. On considère usuellement la vision comme une séquence de traitements de l'image qui permettrait au cerveau de décoder l'information qu'elle comporte. Aujourd'hui, toutefois, cette idée de la vision comme une chaîne de traitement passive de l'information est très sérieusement remise en question. Les sciences cognitives ont vu en particulier émerger des théories dites "actives" de la vision, dans lesquelles l'activité du sujet, au sens moteur du terme, fait partie intégrante du processus de perception visuelle que l'on considère alors comme une boucle liant sensation et action (Gibson, 1986 ; Varela *et al.* 1993 ; O'Regan et Noë, 2002). Dans ce cadre, une nouvelle question se pose : si le traitement de l'information ne constitue plus l'origine même de la perception visuelle, quel est son véritable rôle dans le système visuel ? L'idée est que nous pourrions tenter de répondre à cette question en observant les variations du comportement du sujet lorsque l'on simule ce traitement.

Dans ce chapitre, je poserai les bases nécessaires à la compréhension du manuscrit. Dans une première partie, je ferai un état de l'art sur les prothèses visuelles et sur la question du traitement du signal dans ces prothèses. J'introduirai ensuite les approches actives de la perception visuelle et je présenterai en particulier la théorie des contingences sensori-motrices proposée par Kevin O'Regan. Je présenterai enfin les travaux en lien avec ces théories sur la conception et l'évaluation d'une prothèse visuelle ainsi que sur l'analyse de leur fonctionnement.

## 1.1 Les prothèses visuelles

L'idée de restaurer la vision en transmettant l'information en provenance d'une caméra à un sujet non-voyant date des années soixante. A l'époque, Brindley et Lewin montrent qu'une stimulation

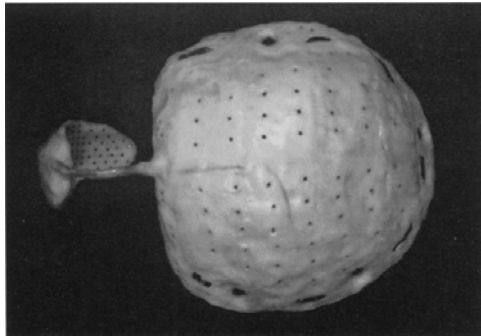


Figure 1.1: le premier implant cortical mis au point par Brindley et Lewin est constitué d'un casque muni de radiorécepteurs (à droite) et d'une interface en contact avec le cortex visuel (à gauche)

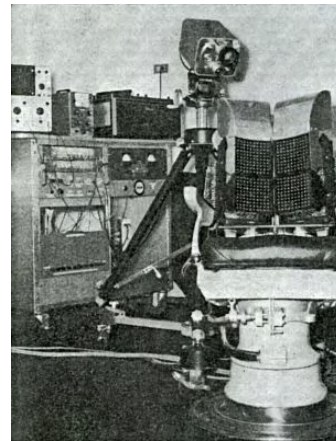


Figure 1.2: le dispositif de Bach-y-Rita, constitué d'une caméra sur pied et d'une matrice de stimulateurs tactiles fixée sur une chaise de dentiste. Le dispositif électronique en arrière-plan convertit l'image en provenance de la caméra en une stimulation tactile

électrique du cortex visuel par une électrode posée à sa surface provoque l'apparition d'un phosphène, c'est-à-dire d'une tache lumineuse, chez un sujet aveugle. Ils montrent également que la stimulation de plusieurs électrodes génère plusieurs phosphènes distincts, et concluent à la faisabilité d'une prothèse visuelle basée sur ce principe, ouvrant la voie aux premiers implants corticaux (Brindley et Lewin, 1968). L'implant de Brindley et Lewin est présenté figure 1.1.

Parallèlement un chercheur du nom de Bach-y-Rita propose de transmettre l'image de la caméra par le biais de la modalité tactile. Son dispositif est constitué d'une caméra et d'une matrice de 20 par 20 pointes métalliques fixées au dos d'une chaise de dentiste (figure 1.2). Ces pointes peuvent être mises en vibration par le passage d'un courant. Un dispositif transcrit de cette manière l'image de la caméra en une stimulation tactile (Bach-y-Rita, 1969). A l'aide de ce dispositif, des sujets aveugles s'avèrent capable de discriminer des patterns visuels simples. La substitution sensorielle était née. La perspective de rendre la vue à un individu aveugle à l'aide de dispositifs optiques et électroniques a constitué un objectif particulièrement enthousiasmant et a suscité un fort engouement. Depuis les travaux pionniers, les dispositifs de suppléance se sont à la fois diversifiés et perfectionnés. Ces travaux ont fait l'objet de plusieurs revues récentes (Veraart *et al.*, 2004 ; Dowling, 2005 ; Merabet *et al.*, 2005 ; Auvray, 2004, Auvray et Myin, 2008)

### 1.1.1 Les implants

Dans une revue de la littérature réalisée en 2005, Dowling recense pas moins de dix-sept projets concernant les prothèses visuelles implantables dans le monde. Ces prothèses diffèrent essentiellement par le lieu d'implantation dans le système visuel.

**Au niveau de la rétine** : dans le cas où le substrat neural n'est pas complètement atteint, et en particulier quand il subsiste suffisamment de cellules ganglionnaires, l'implant au niveau de la rétine est envisageable. L'implant consiste en une matrice d'électrodes placée soit au contact des photorécepteurs (Humayun, 1996, 2003 ; Rizzo *et al.*, 2003) et porte alors le nom d'implant sub-rétinien, soit au contact des cellules ganglionnaires qui forment le nerf optique (Chow et Peachy, 1998 ; Zrenner, 2002) et porte dans ce cas le nom d'implant épi-rétinien.

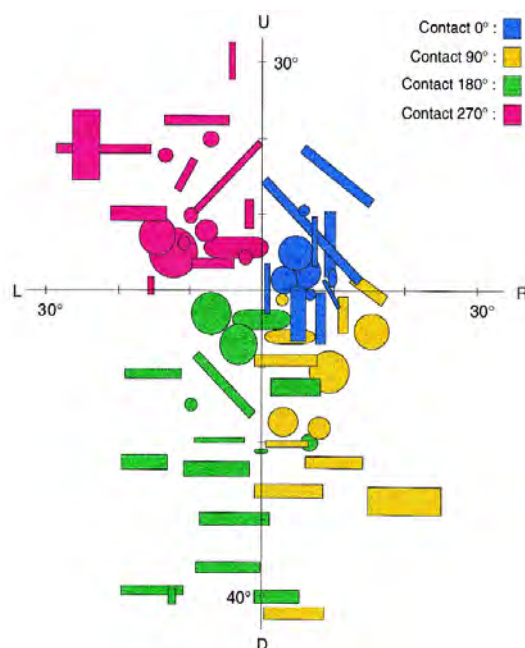


Figure 1.3: phosphènes induits par le manchon optique implanté par l'équipe de Veraart (Veraart, 1998).

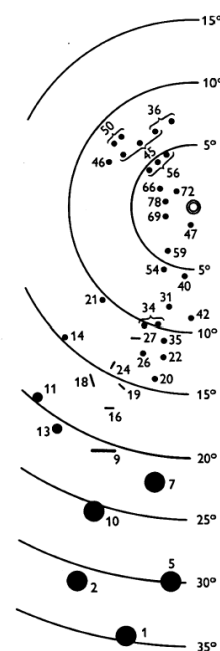


Figure 1.4: phosphènes induits par le dispositif de stimulation corticale de Brindley et Lewin (Brindley et Lewin, 1968)

**Au niveau du nerf optique** : l'équipe de Claude Veraart, en Belgique, a mis au point un manchon qui vient se fixer autour du nerf optique et stimule directement les axones des cellules ganglionnaires (Veraart *et al.*, 2003).

**Au niveau du cortex visuel** : on distingue les implants de surface (Brindley et Lewin, 1968 ; Dobelle *et al.*, 1976 ; Dobelle, 2000) qui sont simplement posés à la surface du cortex, et les implants intracorticaux (Schmidt *et al.*, 1996, Troyk *et al.*, 2003 , 2006 ; Fernandez, 2005) qui consistent en une matrice de micropointes enfoncées dans le cortex.

D'autres projets ont été lancés enfin sur des méthodes de stimulation alternatives, par exemple par le biais de neurotransmetteurs chimiques (Peterman *et al.*, 2003), par le biais d'implants hybrides mêlant support électrique et neurones biologiques (Yagi *et al.*, 2005), ou enfin par des méthodes de stimulation extraoculaire qui ne nécessitent pas de placer les électrodes en contact direct avec la rétine (Kanda *et al.*, 2004, Chowdhury *et al.*, 2005).

Le principe de fonctionnement de tous ces dispositifs est le même. La stimulation électrique du substrat visuel génère ce que l'on appelle des phosphènes (fig. 1.3 et fig. 1.4. Ces phosphènes consistent en l'apparition de taches claires dont la dimension oscille entre la taille d'un grain de riz tenu à bout de bras et la taille d'une demi-allumette. Leur forme et leurs éventuelles colorations varient essentiellement en fonction du type d'implant : les implants les plus périphériques (rétiniens ou sur le nerf optique) produisent en général des phosphènes plus larges, plus colorés et de tailles plus variées que les implants au niveau du cortex. Signalons par ailleurs que ces phosphènes se déplacent avec les mouvements de l'œil. Signalons également qu'il n'a manifestement jamais été possible de produire un phosphène chez un aveugle précoce.

C'est par la suite une caméra, dont le sujet est équipé, qui va piloter l'implant et commander l'apparition de ces phosphènes de manière à restaurer une forme de perception visuelle chez le sujet. Une mise en relation topologique de l'image de la caméra avec la position et éventuellement la

forme des phosphènes est effectuée. On appelle cette opération le "mapping".

En laboratoire, ces dispositifs ont permis aux sujets de distinguer présence et absence de lumière, de localiser des objets, de discriminer la direction du mouvement d'un objet (Humayun, 2003), de discriminer l'orientation d'une lettre (e.g. un L), ainsi que divers symboles ou objets simples (Veraart *et al.*, 2003 ; Yanai *et al.*, 2007), et de lire du braille (Dobelle *et al.*, 1976).

Seul les implants de surface développés par Dobelle ont pu être implantés chroniquement chez les sujets. En 2000, Dobelle publie un article décrivant le cas d'un patient, Jerry, ayant gardé son implant depuis 20 ans. Ce dernier présente une acuité d'environ 20/120 aux tests de reconnaissance de caractères standards. Il rapporte également de manière informelle le fait que le sujet est capable de discriminer trois mannequins : un adulte debout, une femme assise et un enfant de trois ans debout. Il est capable de se déplacer dans une salle en évitant ces mannequins. Il est également capable d'aller chercher un chapeau noir placé à différents endroits sur le mur au fond de la salle, et de revenir le poser sur la tête de l'un des trois mannequins.

Le dispositif de Dobelle est particulièrement intéressant car il a par la suite été implanté chez plusieurs patients. Si aucune de ces implantations n'a fait l'objet par la suite d'une publication dans une revue scientifique, elles ont en revanche eu un certain retentissement médiatique. En 2002, Jens, un fermier du Canada ayant perdu la vue à 18 ans fait le voyage jusqu'au Portugal pour être implanté (Kotler, 2002 ; Gupta et Petersen, 2002). Le dispositif est cette fois constitué de deux matrices d'électrodes, chacune comprenant 64 points de stimulation, une dans chaque hémisphère. Selon Gupta et Petersen, peu après l'opération le sujet est capable de naviguer entre les pièces et de trouver les portes. Le sujet a manifestement également conduit une voiture décapotable sur une portion de parking. Selon Dobelle, les prochaines interfaces pourraient permettre au sujet de conduire sur route. La description que le sujet donne de la tâche semble toutefois relativiser le propos :

"I was able to very carefully drive and look from my left side to my right side, making sure I was between this row of trees on the right and the building on my left"<sup>1</sup>,

ce qui est plus cohérent avec la description qu'il fait de la stimulation :

"I expected them to throw the switch and see everything. but what happened was all I saw was big white splotches on my right side or my left side of my visual field...I was amazed that my visual part of my brain was so out of shape, and could not differentiate between one image and another."<sup>2</sup>

L'institut Dobelle n'a pas autorisé les journalistes à filmer l'utilisation du dispositif au quotidien. Jens rapporte qu'il peut distinguer des mouvements sur l'écran de télévision, et voir les décorations de Noël sur le balcon de son voisin.

En 2005, 16 personnes en tout avaient reçu un implant du Dobelle Institute. Pour Karen Grisdale, implantée en 2003, "[to use the device] is like climbing Mount Everest, and I'm still at sea level"<sup>3</sup>. Il semble donc que ce type d'implant permette d'obtenir des résultats plus qu'anecdotiques au quotidien mais qu'il n'en soit qu'à ses premiers balbutiements. Par ailleurs, les résultats obtenus en dehors de Jerry n'ont pas été documentés.

---

1 Je faisais très attention en conduisant, je regardais à gauche et à droite pour être sûr que je me trouve bien entre la rangée d'arbres à ma droite et le bâtiment à ma gauche

2 Je pensais qu'ils allaient appuyer sur l'interrupteur et que je verrais tout, mais ce que j'ai vu étaient deux grosses flaches lumineuses à gauche et à droite dans mon champ visuel. J'étais surpris de constater que la partie visuelle de mon cerveau soit à ce point sens dessus dessous et que je ne puisse pas différencier une image d'une autre".

3 "Utiliser le dispositif ressemble à grimper le mont Everest, et je me trouve encore au niveau de la mer"



Figure 1.5 : la version portable du TVSS. La caméra est placée au niveau des lunettes, le stimulateur tactile se place au niveau du ventre. D'après Bach-y-Rita, 1972.



Figure 1.6 : visualisation de l'activité des stimulateurs de la matrice tactile. L'image en provenance de la caméra est simplement recopiée sur l'espace bidimensionnel de la matrice tactile. D'après Bach-y-Rita, 1972.

### 1.1.2 La substitution sensorielle

Le principe de la substitution sensorielle est très général. Il consiste à utiliser une modalité de substitution, par exemple le sens tactile ou le sens auditif, pour suppléer ou compléter une autre modalité, en particulier dans le cas où cette dernière est lésée, dans notre cas, la vision. *A priori*, toutes les modalités peuvent être concernées et toutes les combinaisons "modalité de substitution"/"modalité substituée" sont possibles (voir par exemple Bach-y-Rita et Kercel, 2003). Dans le cas présent, nous nous concentrerons sur les dispositifs destinés à se substituer à la modalité visuelle, plus exactement, aux dispositifs destinés à transmettre à un sujet non-voyant un signal en provenance d'une caméra. Nous nommerons ces dispositifs "dispositifs de substitution visuelle".

On distingue principalement deux types de substitution visuelle, en fonction du sens de substitution utilisé.

**La substitution visuo-tactile** : dans ce cas, l'interface consiste en une matrice de stimulateurs mécaniques (des pointes mises en mouvement par des solénoïdes) ou électriques (des électrodes dans lesquelles on fait passer un courant qui provoque une sensation tactile). Les premiers dispositifs de ce type étaient placés au niveau du dos ou du ventre, comme le Tactile Visual Substitution System (TVSS, fig. 1.5) développé par Bach-y-Rita (Bach-y-Rita *et al.*, 1969, Bach-y-Rita, 1972 ; Kaczmarek et Bach-y-Rita, 1995). D'autres interfaces de ce type ont aujourd'hui vu le jour et se placent au niveau de la langue (Bach-y-Rita *et al.*, 1998) ou au niveau du front (Kajimoto *et al.*, 2006).

**La substitution visuo-auditive** : dans ce cas, l'interface consiste en des écouteurs stéréophoniques. Le premier appareil de ce type a été proposé par Fish (1976). Depuis, d'autres dispositifs ont vu le jour, notamment le dispositif The vOICe (Meijer, 1992), le PSVA (Capelle *et al.*, 1998), le dispositif de Cronly-Dillon *et al.* (1999, 2000), et TheVIBE (Auvray *et al.*, 2005).

Le principe de fonctionnement de ces dispositifs est relativement similaire. L'information en provenance de la caméra est "traduite" dans la modalité de substitution. Dans le cas visuo-tactile par exemple, chaque pixel de l'image de la caméra pilote un stimulateur tactile (mécanique ou électrique). L'image au niveau de l'interface est une copie plus ou moins simplifiée de l'image de la



caméra. La figure 1.6 présente l'écran de contrôle du TVSS sur lequel on voit l'état d'activation des stimulateurs de l'interface (1 ou 0, dans ce premier dispositif on ne considérait pas les niveaux de gris).

Dans le cas visuo-auditif, on utilise une analogie entre l'espace bidimensionnel de l'image et deux dimensions, au sens psychophysique, de l'espace sonore, supposées indépendantes (comme par exemple la fréquence d'un son et la position apparente de la source qui le produit, communément appelé panoramique). Nous reviendrons plus en détail sur la question du codage visuo-sonore au chapitre 3.

Le principe de fonctionnement de ces dispositifs laisse peut-être le lecteur perplexe. Quoiqu'il en soit, les résultats obtenus à l'aide de ces dispositifs se sont avérés très encourageants. Dans leur article publié en 1969, Bach-y-Rita et ses collaborateurs rapportent les résultats obtenus chez six patients aveugles, dont cinq sont aveugles de naissance et le sixième depuis l'âge de 4 ans.

Dans une première expérience, ils utilisent des images constituées de bandes noires et blanches parallèles qu'ils présentent au sujet soit horizontalement soit verticalement. Les performances des sujets aveugles sont identiques à celle du groupe contrôle constitué de sujets voyants utilisant l'écran de contrôle.

Dans une deuxième expérience les auteurs utilisent l'image d'un échiquier photographié avec un angle de 70°. La direction de convergence est présentée en haut, en bas, à gauche ou à droite. Encore une fois les sujets sont capables de discriminer les quatre images bien au-delà du hasard. Ils sont toutefois un peu moins performants que les sujets voyants.

Dans une troisième expérience, les sujets devaient reconnaître une lettre présentée sur l'écran. L'opérateur leur indiquait si la réponse était juste. Dans le cas contraire, le sujet devait proposer une autre lettre. Le temps mis pour donner la réponse juste était mesuré. Ce dernier passe de 52 secondes à 10 secondes après trois périodes de test.

Enfin, les auteurs testèrent la capacité des sujets à reconnaître des objets courants. Les premiers essais nécessitaient 5 à 8 minutes d'observation par objet, contre seulement 5 à 20 secondes après 10 heures d'entraînement. Selon les auteurs, les sujets étaient capables de décrire les objets disposés sur une table, en particulier leur position relative en profondeur et cela même si certains objets étaient en occlusion partielle.

Enfin, les auteurs concluent sur l'observation que les sujets décrivent spontanément leur perception comme provenant de devant la caméra et non de la stimulation au niveau de leur dos, ce qu'on appellera par la suite l'externalisation.

Dans un livre publié peu après (Bach-y-Rita, 1972), les auteurs rapportent une quantité d'autres observations individuelles. Ils rapportent notamment le fait que certains sujets sont capables de reconnaître les individus placés devant la caméra, de décrire leur posture, leurs mouvements et leurs caractéristiques individuelles (taille des cheveux, port de lunettes...). Bien que cela semble plus difficile à obtenir, certains sujets s'avèrent même capables de discriminer des photographies de visages. Ils rapportent également la mise en place de comportements réflexes : suite à une manipulation du zoom par l'opérateur à l'insu du sujet, ce dernier a sursauté et a placé ses mains devant lui comme pour éviter une collision.

D'autres études ont par la suite en partie confirmé ces résultats. En laboratoire, les sujets s'avèrent capables de localiser des objets (Jansson, 1983 ; Lenay, 2002 ; Auvray *et al.*, 2007 ; Proulx *et al.*, 2006), de discriminer des formes et des symboles (Bach-y-Rita, 1969 ; Cronly-Dillon, 2000 ; Sampaio *et al.*, 2001 ; Arno *et al.*, 1999, 2001 ; Pollock *et al.*, 2005), de discriminer des objets

(Kaczmarek et Haase, 2003 ; Auvray *et al.*, 2007), de lire (Bliss *et al.*, 1970 ; Loomis, 1974), de réaliser un travail sur une chaîne d'assemblage électronique (Bach-y-Rita *et al.*, 2003) et enfin de conduire un robot mobile dans un environnement simplifié (Segond *et al.*, 2005).

Signalons également que des illusions d'optique ont été reproduites en substitution sensorielle (Guarniero, 1977 ; Renier, 2005).

Enfin, si les dispositifs de substitution ne stimulent pas directement le substrat visuel, des études en imagerie montrent que certaines aires corticales visuelles sont activées par la substitution sensorielle (Arno *et al.*, 2001 ; Pollock *et al.* 2005 ; Poirier *et al.*, 2005 ; Merabet *et al.*, 2009).

On trouvera enfin plusieurs témoignages d'utilisateurs sur le site du concepteur de The vOICe ([www.seeingwithsound.com](http://www.seeingwithsound.com)). Ces derniers sont assez saisissants. Un utilisateur par exemple s'est soudain étonné de percevoir des objets en 3D, un autre se demande pourquoi il se cogne dans les murs et se rend compte alors qu'il a éteint son dispositif pour pouvoir écouter de la musique. Beaucoup d'autres anecdotes encore signalent un véritable potentiel de ce type de dispositif pour une utilisation au quotidien.

## 1.2 Le traitement du signal dans les prothèses visuelles

La piste du traitement du signal pour l'amélioration de l'efficacité des prothèses visuelle a été identifiée dès les premières approches. Dans un article de 1976, Dobelle explique : "Processing and conversion of images for the optimum benefit to the user of a visual prosthesis are areas to which considerable time and effort must be directed in future experiments"<sup>4</sup>

On remarquera en premier lieu la distinction que fait Dobelle entre deux aspects : le traitement de l'image, c'est-à-dire l'ensemble des opérations appliquées à l'image elle-même dans le dispositif avant son envoi vers l'interface, et la conversion de l'image (conversion en impulsions électriques pour les implants, conversion vers le sens de substitution pour la substitution sensorielle) réalisée au niveau de l'interface.

La question de la conversion de l'image au niveau de l'interface est bien sûr très importante dans le développement d'un dispositif, puisqu'il s'agit de savoir *comment* transmettre un signal au sujet. Ce sont par ailleurs les contraintes liées à l'interface qui déterminent le nombre maximal de points de stimulation, ainsi que le nombre de niveaux différents de stimulation à coder. Cette question, en revanche, peut difficilement être considérée d'un point de vue général car elle dépend fortement du type d'interface que l'on considère (implant rétinien, cortical, matrice tactile, etc.) Chaque type d'interface possède ses propres contraintes et ses propres modes de stimulation. Nous aborderons donc cette question plus tard, dans le chapitre 3, dans le cas particulier de la conversion image-son. Nous montrerons en particulier comment les outils du traitement du signal et de la théorie de l'information peuvent être mis à profit pour améliorer la "bande passante" de l'interface.

La question du traitement de l'image est en revanche commune à toutes les prothèses visuelles. Il s'agit de savoir *quelle image doit être envoyée* à l'interface.

Dans la pratique aujourd'hui, le prétraitement de l'image dans les prothèses visuelles avant son envoi à l'interface se résume en général à deux choses : la pixelisation de l'image, c'est à dire sa réduction du format vidéo au format de l'interface, auquel s'ajoute dans le cas particulier des implants visuels, le "mapping", qui consiste à la mise en relation de la position du pixel dans l'image avec celle du phosphène dans le champ visuel du sujet et que nous n'aborderons pas ici.

---

4 " Le traitement et la conversion de l'image pour un bénéfice optimal de l'utilisateur d'une prothèse visuelle sont des domaines dans lesquels nous devons consacrer beaucoup de temps dans les travaux futurs".



Figure 1.7: simulation de la pixelisation dans une prothèse visuelle. A gauche l'image d'origine, à droite l'image pixelisée présentée via des lunettes de réalité virtuelle. D'après Chen *et al.* 2004

Concernant la pixelisation de l'image, la solution consiste en général à sous-échantillonner directement l'image de la caméra, c'est-à-dire à "sauter" une grande partie des pixels (Bach-y-Rita *et al.*, 1969 ; Meijer, 1992), ou à effectuer des opérations de moyennes sur des zones de l'image (Yanai *et al.*, 2007). La méthode de sous-échantillonnage n'est par ailleurs pas toujours précisée (Dobelle *et al.*, 1976 ; Veraart *et al.*, 2003). Mentionnons toutefois une méthode de sous-échantillonnage spatialement variante utilisée dans le PSVA (Prothèse de Substitution de la Vision par l'Audition) qui consiste à faire une moyenne sur des zones plus petites au centre de l'image que dans la périphérie (Capelle *et al.*, 1998). L'intérêt de ce type d'échantillonnage par rapport à un échantillonnage uniforme n'a toutefois pas été évalué. Signalons également le cas du dispositif TheVIBE (Auvray *et al.*, 2005), dans lequel la méthode d'échantillonnage est entièrement configurable par l'utilisateur. Nous reviendrons en détail sur cela au chapitre 3.

De nombreuses études sur la question de la pixelisation ont en revanche été effectuées à l'aide de dispositifs destinés à simuler l'utilisation d'une prothèse visuelle chez des sujets voyants. On utilise pour cela un casque de réalité virtuelle dans lequel l'image pixellisée est projetée (fig. 1.7). On étudie ainsi l'incidence de cette pixelisation sur les performances du sujet, ainsi que différentes méthodes de sous échantillonnages (Cha *et al.*, 1992a, 1992b ; Smets et Overbeeke, 1995 ; Hayes *et al.*, 2003 ; Thompson, 2003 ; Chen *et al.*, 2004, 2006 ; Sommerhalder *et al.*, 2004 ; Pérez-Fornos *et al.*, 2005 ; Vurro *et al.*, 2006 ; Hallum *et al.*, 2005, 2008 ; Dagnelie 2007). Les auteurs font en général varier le nombre de pixels de l'image (c'est à dire le nombre de points de l'interface), leur organisation (la manière dont ils sont agencés, par exemple en grille carrée ou hexagonale), la forme et la taille de la région de l'image que ce point va représenter (que l'on appelle en générale un "champ récepteur"), et enfin le nombre de niveaux de gris disponibles (le nombre de niveaux de stimulation du point). Pérez-Fornos étudie également l'ajout d'une intégration temporelle dans le calcul du phosphène. L'étude de Cha sur la mobilité et celle de Sommerhalder s'intéressent par ailleurs à la taille du champ de la caméra.

Globalement ces études ont montré que le nombre limité de points typique pour une prothèse visuelle, de l'ordre de 100 à 1000 points, et celui de niveaux de gris, de 4 à 16 en général, restent compatibles avec des tâches de mobilité et de lecture. Le nombre de points reste néanmoins le facteur principal quant aux performances des sujets. L'organisation et la forme des "champs récepteurs" restent des éléments secondaires. Ce type d'étude n'a toutefois pas à ma connaissance été mené en pratique sur une prothèse visuelle et la question de savoir si les résultats en simulation sont équivalents au cas réel reste entière. Par ailleurs, aucune étude n'a été menée sur l'intérêt d'un

échantillonnage spatialement variant. Nous reviendrons en détail sur ces questions dans la suite du propos.

Dans certains dispositifs récents toutefois, une étape de prétraitement de type "détection de contour" a été ajoutée, à l'aide d'outils classiques en traitement du signal, en l'occurrence un filtre de Sobel (Dobelle, 2000), et un filtre de type Laplacien (Kajimoto, 2006). Les auteurs n'ont toutefois pas cherché à évaluer l'apport de ce type de filtrage. Signalons néanmoins un travail préliminaire sur l'intérêt de ce type de filtre pour la reconnaissance d'image réalisé par Laloyaux et Schmitt (2003).

Si la phase de prétraitement de l'image est relativement sommaire dans les dispositifs actuels, de nombreuses études de faisabilité ont été réalisées et de nombreux logiciels destinés à prétraiter l'image de la caméra ont vu le jour. Si certains sont basés sur les outils classiques du traitement de l'image (par exemple Liu *et al.*, 2005), la plupart s'inspirent directement du système visuel humain (Fernandez *et al.* 2005 ; Eckmiller *et al.*, 2005 ; Zaghloul et Boahen, 2006 ; Asher *et al.*, 2007). Ces approches sont dites biomimétiques.

Les approches biomimétiques se concentrent principalement sur les premières étapes de traitement dans le système visuel : celles qui ont lieu avant l'arrivée du signal au cortex visuel primaire, notamment au niveau de la rétine. Elles visent en général à reproduire les quatre étapes importantes de cette chaîne de traitement :

- 1- L'échantillonnage induit par la distribution spatialement variante des photorécepteurs au niveau de la rétine
- 2- Les adaptations aux conditions d'éclairement et de contraste au niveau de la rétine
- 3- La séparation du signal en plusieurs voies, en fonction de son contenu spatial et temporel, au niveau du nerf optique
- 4- Les propriétés géométriques induites par la projection du tractus optique sur le cortex visuel

Ces quatre étapes seront développées dans le prochain chapitre où nous proposerons un modèle et une simulation de cette chaîne rétinocorticale.

L'enjeu de ces approches biomimétiques est double. Tout d'abord, elles devraient permettre d'améliorer l'efficacité d'un prothèse visuelle, dans la mesure où elles copient une partie du fonctionnement du système visuel. En particulier, la chaîne rétinocorticale est la partie la plus périphérique du système visuel et est à ce titre très spécialisée. Son rôle est vraisemblablement de "mâcher le travail" du sujet, par un formatage adéquat du signal.

Par ailleurs, ce type d'approche présente un enjeu sur le plan de la recherche fondamentale. Elle soulève notamment la question de savoir s'il va suffire d'envoyer le bon signal au bon endroit pour restaurer une perception visuelle, à compter bien sûr que cela soit possible, voire même que cela ait un sens. En effet la question de savoir s'il suffit de transmettre un signal, de provoquer une sensation, pour générer une perception fait aujourd'hui l'objet de nombreuses controverses. Ces controverses, d'ordre philosophique, interrogent la définition même de la notion de perception. Elles posent la question suivante : "Qu'est-ce que voir ? ". Leur enjeu dans le domaine de la conception des prothèses visuelles est par ailleurs évident.

## 1.3 Qu'est-ce que voir : une question très actuelle

A l'époque d'Euclide (300 av. J.C.), on considérait que c'est le regard lui-même qui éclairait le monde comme une main qui se poserait sur les objets alentours. Cette perspective a bien sûr

beaucoup évolué. On sait en particulier que ce n'est pas notre regard qui éclaire les objets mais des sources extérieures de lumière, que cette lumière est réfléchi par les objets et qu'elle entre ensuite dans nos yeux. Cette idée a amené à penser que voir consiste en l'analyse par notre cerveau de cet influx visuel. Elle est toutefois aujourd'hui remise en question, et cette analogie avec une main qui explorerait le monde retrouve une place avec un sens nouveau : la vision serait une activité d'exploration et non l'analyse passive d'une sensation.

### 1.3.1 Qu'est-ce que voir : les approches actives de la perception visuelle

On considère en général la perception visuelle comme une série d'étapes de traitement grâce auxquelles le cerveau extrairait, pas à pas, la substantifique information présente dans le signal émis par les photorécepteurs. Cette présentation somme toute didactique du fonctionnement du système visuel n'est toutefois que très partielle, car s'il est vrai que le signal issu des photorécepteurs se propage le long des voies du système visuel, cela n'explique en rien comment le cerveau peut comprendre cette information. On appelle cette question le "fossé sémantique" : comment passe-t-on du signal à l'information?

Le présupposé à ce sujet consiste en général à faire appel à la notion de représentation. Une représentation mentale serait une conformation physique du cerveau qui aurait pour propriété de faire référence à un objet du monde extérieur. Par exemple, la notion de chaise pourrait correspondre à l'activation d'un neurone particulier, ou à un pattern d'activation particulier de plusieurs neurones. Le traitement dans le système visuel aurait alors pour fonction d'activer ces représentations et de construire un lien entre les différents "objets" représentés, lien qui correspondrait à un lien faisant sens dans le "monde extérieur".

Cette position est issue de ce que Varela appelle "l'hypothèse cognitiviste" (Varela *et al.*, 1993). Elle consiste à définir la cognition comme le traitement de l'information en tant que computation symbolique, c'est à dire une manipulation de symboles basée sur des règles. Le modèle de la vision proposé par Marr (1982) est essentiellement basé sur cette idée. Le système visuel extrait des primitives de bas niveau (des formes géométriques de base) qui sont ensuite assemblées en objets eux-mêmes agencés les uns par rapport aux autres pour former la scène visuelle.

La position cognitiviste est aujourd'hui fortement remise en question en Sciences Cognitives. Le lecteur trouvera dans Varela une analyse détaillée à ce sujet. En particulier, la position cognitiviste implique la notion d'un monde "pré-défini" et notre intelligence consisterait en une adéquation entre ce monde pré-défini et un ensemble de symboles et de règles : "Pour les écoles représentationnistes, une entité cognitive est pour l'essentiel parachutée dans un monde pré-existant. Cette entité n'y survivra qu'à la condition de posséder une carte et d'apprendre à agir en fonction de celle-ci" (Varela, 1989, p 102).

Par ailleurs, la perspective cognitiviste s'applique mal au cas de la substitution sensorielle. On peut légitimement se demander par exemple comment un son ou une stimulation tactile pourrait activer une représentation de nature visuelle. On imagine en tout cas que cela devrait impliquer des mécanismes complexes de réorganisation du cerveau, ce qui est peu compatible avec la rapidité avec laquelle les sujets s'approprient le dispositif. A moins de postuler l'existence préalable de représentations dédiées à la substitution sensorielle.

Les théories actives de la vision permettent de proposer une autre approche du processus de perception visuelle. On trouvera leur expression récente chez trois auteurs en particulier : Gibson (1986), Varéla (1993), et O'Regan et Noë (2001). Le point commun de ces théories est de séparer la

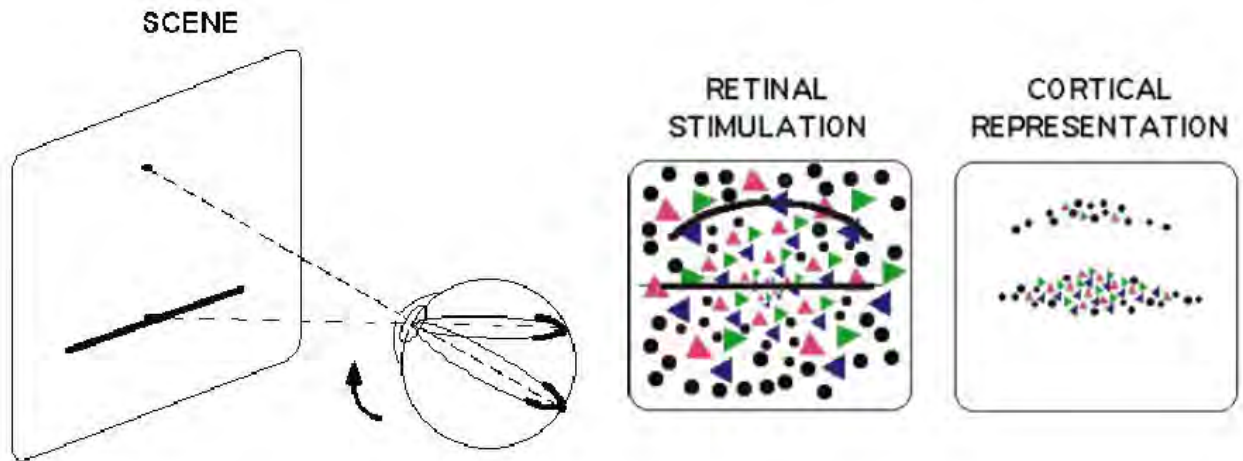


Figure 1.8: Illustration de l'expérience de pensée de O'Regan et Noë sur la perception d'une ligne droite. Modifié, d'après O'Regan et Noë, 2001

notion de sensation (le signal) et la notion de perception (l'information), et d'introduire l'activité du sujet comme un élément fondamental dans le processus perceptif, au même titre que la sensation. On trouvera dans Mossio et Taraborelli (2008) et Taraborelli et Mossio (2008) une analyse fine des points communs et des points de divergence de ces trois approches. Je me contenterai pour ma part de présenter l'approche proposée par O'Regan et Noë, sur laquelle nous nous sommes basés pour travailler.

### 1.3.2 La théorie des contingences sensori-motrices

La théorie proposée par O'Regan et Noë (2001) repose sur la notion de "contingence sensori-motrice". Une contingence sensori-motrice est un lien particulier qui s'établit entre une variation liée à une action et une variation de nature sensorielle. Un exemple très simple de contingence sensorimotrice au niveau de la vision : quand je ferme mes paupières, je passe de la lumière à l'obscurité. Un autre exemple, auquel nous n'avons pas accès au sens phénoménologique du terme : lorsque je bouge mes yeux dans un sens, la stimulation sur la rétine se déplace dans le sens opposé.

Par ailleurs, O'Regan et Noë définissent deux types de contingences sensori-motrices : celles liées à l'appareil de mesure (c'est le cas de la contingence présentée plus haut) et celles liées aux propriétés mêmes du "monde visuel" (compris non en tant que monde réel préexistant mais en tant que phénomène en interaction avec le dispositif de mesure visuel). Les contingences sensori-motrices liées au monde visuel sont régies par exemple par les notions de perspective, de parallaxe et d'occlusion. Selon O'Regan et Noë, la vision consisterait en une activité d'exploration du monde médiée par la connaissance qu'a le sujet des lois régissant ces contingences sensori-motrices. Pour éclairer leur propos les auteurs proposent un exemple particulièrement didactique : celui d'une ligne droite. Ils proposent en effet de considérer le cas d'un observateur regardant une ligne droite. Cette ligne droite produit une activation de certaines cellules de la rétine et du cortex visuel. Cette activation est schématisée figure 1.8. Supposons maintenant que le regard de l'observateur pointe au dessus de la ligne. L'image formée au niveau des photorécepteurs de cette ligne subit alors une déformation qui se répercute sur l'image corticale. Si l'on suppose une approche passive de la perception, le système visuel doit être capable de reconnaître en ces deux patterns différents une même ligne droite.

Dans l'approche sensori-motrice, O'Regan et Noë proposent de considérer qu'une ligne droite est un objet invariant par translation du regard. En effet, si je translate mon regard le long d'une ligne droite, la stimulation au niveau de la rétine reste inchangée, le pattern d'activation cellulaires qu'elle produit reste de ce fait également inchangé, et ce quel que soit la nature en pratique de ce pattern d'activation. L'information ne réside plus ici dans le pattern d'activation lui-même mais dans le fait que ce pattern soit invariant lors d'une action motrice donnée.

Les conséquences de ce changement de point de vue sur la nature du processus visuel lui-même et sur la redéfinition de ce que nous appelons le monde visuel entraîne de nombreuses conséquences. L'article de O'Regan et Noë propose à ce titre une analyse de nombreux phénomènes sous ce nouvel angle de vue et montre comment il permet d'apporter un éclairage nouveau sur de nombreuses questions, comme par exemple le fait que nous ne percevions pas notre tache aveugle : ce "trou" de la taille d'un citron à bout de bras dans notre champ visuel qui est lié au passage du nerf optique hors de l'œil, produisant une zone où il n'y a pas de cellules photosensibles, ainsi qu'un nouveau regard sur la notion de "modalité" perceptive. Selon O'Regan et Noë, une modalité perceptive consiste en un répertoire particulier de contingences sensori-motrices. Voir consiste ainsi non pas dans le fait de recevoir une information sensorielle en provenance des yeux mais dans le fait d'utiliser un dispositif sensoriel (artificiel ou naturel) utilisant des répertoires de contingence sensori-motrices de nature visuelle. Cette idée est bien sûr très intéressante dans le cadre de la conception d'une prothèse. La théorie des contingences sensori-motrices a ainsi donné lieu à de nombreux travaux sur le fonctionnement des dispositifs de substitution et a contribué à l'émergence d'une thématique de recherche : l'ergonomie sensori-motrice des systèmes de perception artificiels et naturels (Gapenne et Gaussier, 2005).

## 1.4 Approche sensori-motrice et prothèses visuelles

Dans le cadre de la théorie des contingences sensori-motrice, une prothèse visuelle ne peut plus être considérée seulement comme un dispositif destiné à transmettre un signal. Selon Lenay *et al.* (2000), elle doit être considérée, et même redéfinie, comme un dispositif de couplage sensori-moteur (fig. 1.9).

Cette nouvelle approche a des incidences fortes non seulement sur la manière de concevoir une prothèse visuelle, mais aussi sur la manière de l'évaluer et finalement sur la manière de comprendre comment le sujet s'en sert. Ce dernier point permet en particulier de définir une nouvelle méthodologie pour étudier les systèmes perceptifs naturels et artificiels.

### 1.4.1 Ergonomie sensori-motrice d'une prothèse visuelle

Dans une approche "classique", les questions de conception sont associées uniquement à la transmission du signal. Il s'agit donc essentiellement de savoir *comment* transmettre le signal. L'interface est à ce titre le point de conception le plus essentiel. Concernant le signal en lui-même, l'approche va être essentiellement quantitative : quelle "quantité" de signal peut-on (au regard de la capacité de l'interface) ou doit-on (au regard de la charge cognitive impliquée chez le sujet) transmettre ?

Dans un cadre classique par ailleurs, l'accent va être mis sur la fidélité du signal transmis à partir de l'image capturée par la caméra. En effet, si je veux que mon cerveau puisse "reconnaître" l'information que je lui "envoie", elle se doit d'être le plus fidèle possible au signal que ce dernier "reçoit" usuellement. Ceci nécessite en particulier que l'interface soit parfaitement "calibrée" de manière à transmettre fidèlement le signal qu'on lui envoie.

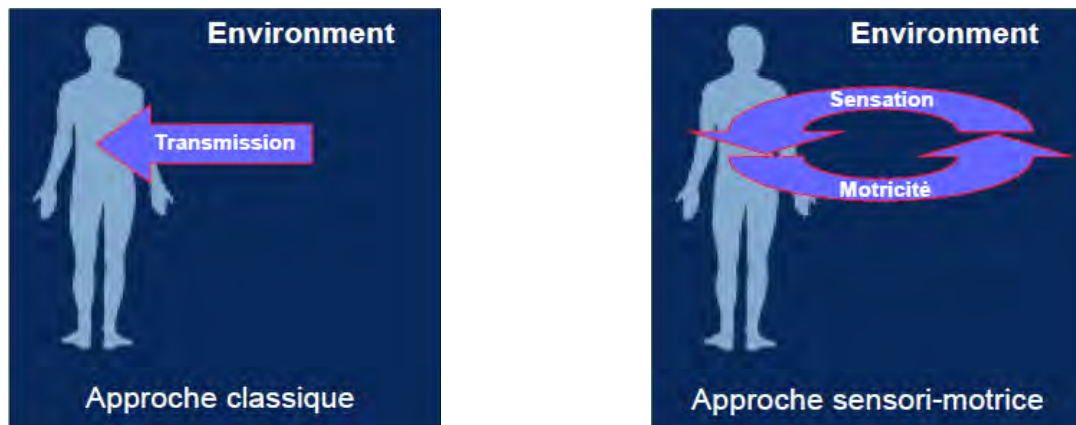


Figure 1.9: illustration des deux approches concernant la nature d'un dispositif de perception.

Dans le cadre classique, les questions de conception concerneront donc essentiellement les points suivants :

- 1 – Le capteur : notamment sa sensibilité et ses spécifications (quantification, résolution, etc.)
- 2 – La qualité de l'image : notamment en terme de luminance et de contraste
- 3 – La qualité de l'interface : notamment en terme de fiabilité et de calibration

Le fait de redéfinir une prothèse visuelle comme un dispositif de couplage étend considérablement le champ de l'ergonomie d'une telle prothèse. Par ailleurs, si les questions associées dans le cadre classique à la transmission du signal restent valides, elles prennent un sens nouveau dans un cadre sensori-moteur.

Dans un cadre sensori-moteur, les questions relatives à la conception du dispositif concernent tout naturellement trois points : la sensation, la motricité et le couplage entre sensation et motricité.

### **La sensation :**

La question de la sensation dans un dispositif de couplage sensorimoteur s'apparente à la question de la transmission dans une approche classique. Elle regroupe donc les points énoncés précédemment à ce sujet.

Il existe néanmoins une différence importante entre la question de la sensation dans un dispositif de couplage sensori-moteur et la question de la transmission dans une approche classique. Dans une approche sensori-motrice en particulier, la perception est issue du lien entre les *variations* de la sensation et les variations motrices. De ce fait, on peut supposer que la valeur nominale du signal a peu d'importance au regard des variations de ce signal. Dans ce cadre, la qualité première d'une interface est donc de permettre une bonne discrimination des variations du signal, et non de reproduire de manière fiable une sensation donnée (au sens psychophysique du terme). Ceci en particulier donne un tout autre regard sur les problèmes d'habituation du canal sensoriel au signal (ce phénomène est par exemple présenté dans Kaczmarek, 1995).

Nous verrons plus en détail dans la suite du manuscrit comment cette question peut influencer la conception du dispositif, et notamment le choix des "échelles psychophysiques" utilisées pour générer la sensation.



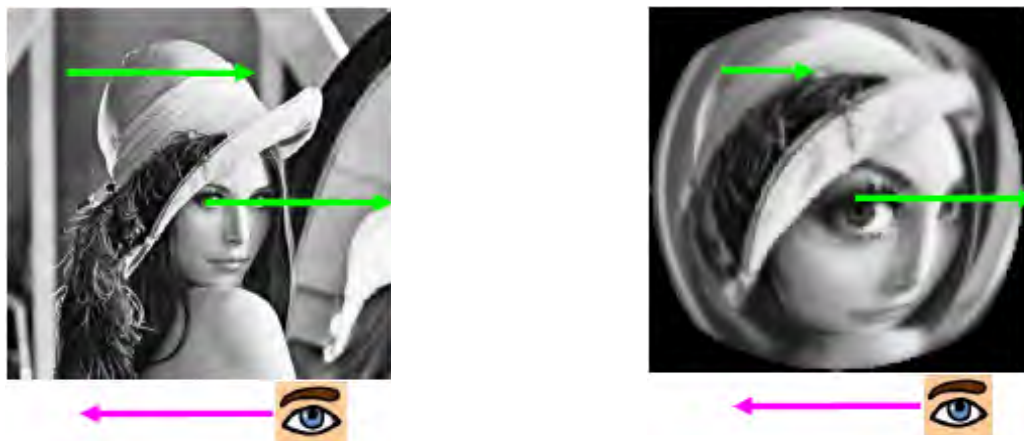


Figure 1.10: modification du couplage sensorimoteur induite par un échantillonnage spatialement variant de l'image. Les flèches vertes représentent la vitesse de déplacement de deux points de l'image, en périphérie et au centre, lors d'une translation de l'œil (ou de la caméra). Dans le cas d'un échantillonnage uniforme (à gauche), la vitesse de déplacement est la même en tous les points. Dans le cas d'un échantillonnage spatialement variant (à droite), elle varie en fonction de l'excentricité du point.

### La motricité :

La question de la prise en compte de l'ergonomie du dispositif en terme moteur est soulignée notamment dans Lenay *et al.*, 2000. Dans un cadre sensori-moteur, deux dispositifs ne sont pas équivalents s'ils n'offrent pas le même espace des possibles en terme moteur. Un dispositif constitué d'une caméra fixe posée sur un trépied est très différente d'un dispositif où la caméra est posée par exemple sur le front. Ce point a par ailleurs amplement été démontré puisque dans le premier cas, les capacités du sujet s'avèrent beaucoup plus limitées que dans le deuxième cas (Bach-y-Rita, 1972).

### Le couplage entre sensation et motricité :

Comme nous l'avons dit, selon l'approche sensori-motrice, le processus perceptif entier repose sur le couplage entre variation de sensations et variations motrice.

La forme concrète de ce couplage est pour partie inhérente au dispositif de perception : ce sont les lois de contingences sensorimotrices associées à l'outil. Considérons à titre didactique le cas d'un échantillonnage spatialement variant du type de celui en cours dans le PSVA. Cet échantillonnage conduit à une magnification au centre de l'image, un peu comme un effet de loupe (figure 1.10 à droite). Ce type de modification introduit des changements au niveau du couplage sensori-moteur. Dans le cas présent, on montre le cas d'une translation de l'œil ou de la caméra. Dans le cas d'un échantillonnage uniforme, la vitesse de déplacement de l'image sur la rétine ne dépend pas du point de l'image considéré. Dans le cas de l'échantillonnage spatialement variant proposé ici, elle décroît avec l'excentricité : le lien entre les variations motrices et les variations de sensation a effectivement changé.

Intuitivement, on peut proposer deux critères pour caractériser le couplage induit par l'appareil : sa richesse et sa qualité. La richesse du couplage tient au fait que le lien entre motricité et sensation est particulièrement ténu et qu'une variation motrice provoque facilement une variation de sensation. La qualité du couplage consiste en une adéquation entre les caractéristiques du couplage, opérationnalisées notamment par l'existence de certaines invariances sensori-motrices qui font sens par rapport à l'activité du sujet. Par exemple dans la marche, une invariance liée à un effet de zoom



Figure 1.11: le dispositif minimaliste Tactos II utilisé dans l'expérience de Lenay. D'après Lenay (2002)



Figure 1.12: mesure de la trajectoire du stylet tactile lors de l'affichage de la lettre S. D'après Lenay (2002)

est vraisemblablement plus intéressante qu'une invariance liée à une translation de l'image. Nous reviendrons sur cette proposition dans la suite du propos.

### 1.4.2 Immersion dans un dispositif de substitution : un cadre pour l'évaluation

Les travaux sur les prothèses visuelles dans le cadre d'une théorie sensori-motrice de la perception n'ont pas seulement permis d'identifier des points de conception jusqu'alors peu considérés, mais également d'étudier la genèse de la perception via une prothèse visuelle. Ces travaux ont permis en particulier de définir des étapes d'immersion dans le dispositif (Auvray, 2004).

Gapenne et Gaussier soulignent à juste titre l'importance de cette démarche dans le cadre de l'évaluation d'un dispositif. En effet, il est vraisemblable que certaines tâches comportementales ne soient accessibles au sujet qu'à partir d'un certain stade d'immersion (comme par exemple éprouver une émotion en regardant la photo d'un être cher). La méconnaissance de ces stades pourrait naturellement conduire à considérer à tort un dispositif comme inutile ou inefficace. Par ailleurs, la connaissance de ces stades est susceptible de permettre la conception de protocoles d'apprentissage efficaces.

### 1.4.3 Une méthodologie d'étude des processus perceptifs

Il existe un écueil épistémologique fort dans l'idée de rendre compte des mécanismes perceptifs à partir des connaissances que nous avons constituées précisément à l'aide de ces mêmes mécanismes. Le cas de la perception de l'espace est particulièrement significatif à ce sujet. Par exemple, je peux postuler l'existence d'un espace tridimensionnel dans lequel je me situe, je peux ensuite en étudier les propriétés et en déduire une loi qui me donne la taille objective d'un objet en fonction de sa distance. Doit-on pour autant considérer que c'est par ce même calcul que je détermine la taille d'un objet et qu'en vertu de ce calcul j'éprouve un phénomène de constance perceptive, c'est-à-dire que sa taille me paraît identique quelque soit sa distance ?

Selon de nombreux philosophes, comme par exemple Merleau-Ponty, la réponse est non. Doit-on pour autant abandonner l'idée d'objectiver le fonctionnement de notre système perceptif ? Selon

Charles Lenay, la réponse est également non. Il propose à ce sujet une méthode qui permet de faire le lien entre l'expérience à la première personne (la manière dont le sujet réalise la tâche "de l'intérieur") et l'objectivité caractéristique de l'analyse en troisième personne (celle d'un observateur analysant le comportement du sujet). Pour éclairer le propos, penchons-nous sur une expérience proposée par Lenay (2002).

Dans cette expérience, le sujet est muni d'un dispositif de substitution minimaliste, constitué d'un stylet pour tablette graphique (le capteur) relié à un ordinateur pilotant un unique stimulateur tactile (l'effecteur, constitué d'un picot tactile stimulant le doigt) en mode tout ou rien (figure 1.11). La tâche du sujet consiste à explorer, à l'aide du stylet tactile, l'espace 2D constitué par une tablette graphique. Grâce à l'ordinateur, une lettre est affichée virtuellement sur la tablette graphique : lorsque le stylet croise le corps de cette lettre, le picot se met à vibrer, autrement il reste silencieux.

La tâche du sujet consiste à identifier la lettre affichée virtuellement. Le déplacement du stylet sur la tablette graphique est enregistré pendant chaque essai (fig. 1.12).

Après avoir montré que le taux de reconnaissance des sujets était supérieur au hasard (analyse en troisième personne), Lenay propose une description en première personne de l'exploration d'une ligne droite que je vais reproduire partiellement ici :

"Je suis dans le noir, les yeux bandés, la main droite tenant un stylet, la main gauche posée sur un petit système de stimulation tactile. Je commence par balayer l'espace de la tablette à la recherche d'une variation sensorielle. Dès que je l'ai trouvée, j'inverse et diminue mon geste pour converger sur le centre de stimulation par un petit balayage. [...] Cependant, je rencontre plusieurs fois la stimulation pour des successions de mouvements décalés. [...] Je maintiens mon petit balayage en le combinant à un mouvement général plus lent.[...] Avec mon petit balayage, non seulement je constitue une présence sous-jacente aux stimulations, mais pratiquement, quand je ne retrouve pas la stimulation je peux avoir une idée sur la direction à suivre pour la retrouver.[...] Si je maîtrise bien un savoir-faire qui me permet de maintenir une variation régulière des retours sensoriels, alors je constitue un objet "Ligne droite *continue*" présent dans l'espace de mes actions."

La méthode de Lenay consiste donc dans un premier temps à identifier les *stratégies* du sujet.

Le pas essentiel franchi par Charles Lenay est de considérer que ces stratégies peuvent alors être formalisées et faire l'objet *a posteriori* d'une analyse en troisième personne. La figure 1.13 montre la corrélation entre la vitesse du stylet et les phases d'activation du stimulateur. Dans le cas présent, d'un sujet réussissant la tâche, on observe une corrélation forte entre les pics de vitesse du stylet et le déclenchement du stimulateur, ce qui montre que le sujet anticipe la position de la ligne. Dans le cas d'un sujet ne réussissant pas cette tâche, on n'observe pas cette corrélation. On a donc ici une méthode permettant de prédire si, oui ou non, le sujet perçoit une ligne droite. C'est à ce titre que l'on peut considérer avoir objectivé la perception d'une ligne droite avec ce dispositif.

Lenay a baptisé cette méthode "méthode des répondants", en ce sens que l'analyse en première personne et son objectivisation en troisième personne se répondent mutuellement.

Appliquée à la question du rôle du traitement du signal dans le système visuel, cette méthode est d'un intérêt fondamental. En effet, dans ce cadre épistémologique, si j'observe des différences entre les stratégies des sujets corrélées au traitement de l'image dans le dispositif, je me trouve à même de comprendre quel rôle effectif joue le traitement du signal dans le dispositif. Et dans la mesure où le traitement du signal que j'utilise est directement inspiré du système visuel, je suis à même d'identifier des pistes concernant le rôle effectif du traitement du signal dans le système visuel.

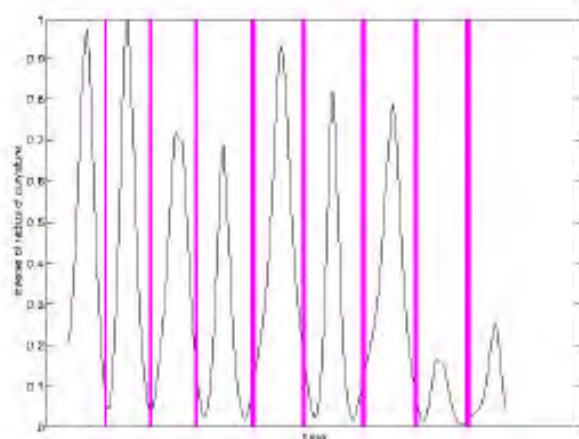


Figure 1.13: corrélation temporelle entre la vitesse de déplacement du stylet (courbe en noir) et l'activation du stimulateur tactile (barres horizontales). Lorsque le sujet réussit la tâche, le déclenchement du stimulateur tactile correspond aux pics de vitesse du stylet. D'après Lenay, 2002.

Il est vraisemblable par ailleurs que cette meilleure compréhension du rôle effectif du traitement du signal dans le dispositif permette en retour d'en améliorer encore l'efficacité. Si l'on acquière en effet une meilleure compréhension du véritable rôle fonctionnel du traitement du signal dans le dispositif, nous serons mieux à même de l'adapter à la prothèse elle-même et en particulier d'en régler les différents paramètres.

Maintenant que le cadre théorique et les enjeux subséquents sont établis, voyons plus en détail en quoi consiste la modélisation de la chaîne rétino-corticale que nous nous proposons d'implémenter dans un dispositif de substitution sensorielle.

## Références

- Arno, P.; Capelle, C.; Wanet-Defalque, M. C.; Catalan-Ahumada, M. & Veraart, C. (1999), 'Auditory coding of visual patterns for the blind.', *Perception* **28**(8), 1013--1029.
- Arno, P.; Volder, A. G. D.; Vanlierde, A.; Wanet-Defalque, M. C.; Streel, E.; Robert, A.; Sanabria-Bohórquez, S. & Veraart, C. (2001), 'Occipital activation by pattern recognition in the early blind using auditory substitution for vision.', *Neuroimage* **13**(4), 632--645.
- Asher, A.; Segal, W. A.; Baccus, S. A.; Yaroslavsky, L. P. & Palanker, D. V. (2007), 'Image processing for a high-resolution optoelectronic retinal prosthesis.', *IEEE Trans Biomed Eng* **54**(6 Pt 1), 993--1004.
- Auvray, M. (2004), 'Immersion et perception spatiale : L'exemple des dispositifs de substitution sensorielle', PhD thesis, Ehess, Paris.
- Auvray, M.; Hanneton, S.; Lenay, C. & O'Regan, K. (2005), 'There is something out there: distal attribution in sensory substitution, twenty years later', *Journal of Integrative Neuroscience* **4**(4), 505.
- Auvray, M.; Hanneton, S. & O'Regan, J. K. (2007), 'Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with 'The vOICe'', *Perception* **36**, 416-430.

- Auvray, M. & Myin, E. (2009), 'Perception with compensatory devices : from sensory substitution to sensorimotor extension', *Cognitive Science* **in press**.
- Bach-y-Rita, P. (1972), *Brain mechanisms in sensory substitution*, Academic Press Inc.
- Bach-y-Rita, P.; Collins, C. C.; Saunders, F. A.; White, B. & Scadden, L. (1969), 'Vision substitution by tactile image projection.', *Nature* **221**(5184), 963--964.
- Bach-y-Rita, P.; Kaczmarek, K. A.; Tyler, M. E. & Garcia-Lara, J. (1998), 'Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note.', *J Rehabil Res Dev* **35**(4), 427--430.
- Bach-y-Rita, P. & Kercel, S. W. (2003), 'Sensory substitution and the human-machine interface.', *Trends Cogn Sci* **7**(12), 541--546.
- Bach-y-Rita, P.; Tyler, M. E. & Kaczmarek, K. A. (2003), 'Seeing with the brain', *International journal of human-computer interaction* **15**(2), 285-295.
- Bliss, J.; Katcher, M.; Rogers, C. & Shepard, R. (1970), 'Optical-to-tactile image conversion for the blind', *IEEE Transactions on Man-Machine Systems* **11**(1), 58--65.
- Brindley, G. S. & Lewin, W. S. (1968), 'The sensations produced by electrical stimulation of the visual cortex', *J. Physiol.* **196**, 479-493.
- Capelle, C.; Trullemans, C.; Arno, P. & Veraart, C. (1998), 'A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution', *Biomedical Engineering, IEEE Transactions on* **45**(10), 1279--1293.
- Cha, K.; Horch, K. W. & Normann, R. A. (1992), 'Mobility performance with a pixelized vision system.', *Vision Res* **32**(7), 1367--1372.
- Cha, K.; Horch, K. W.; Normann, R. A. & Boman, D. K. (1992), 'Reading speed with a pixelized vision system.', *J Opt Soc Am A* **9**(5), 673--677.
- Chen, S.; Hallum, L.; Suaning, G. & Lovell, N. (2006), Psychophysics of prosthetic vision: I. Visual scanning and visual acuity, in 'Conf Proc IEEE Eng Med Biol Soc'.
- Chen, S.; Lovell, N. & Suaning, G. (2004), 'Effect on prosthetic vision visual acuity by filtering schemes, filter cut-off frequency and phosphene matrix: a virtual reality simulation.', *Conf Proc IEEE Eng Med Biol Soc* **6**, 4201--4204.
- Chow, A. & Peachy, N. (1998), 'The Subretinal Microphotodiode Array Retinal Prothesis', *Ophthalmic research* **30**(3), 195--196.
- Chowdhury, V.; Morley, J. W. & Coroneo, M. T. (2005), 'Feasibility of extraocular stimulation for a retinal prosthesis.', *Can J Ophthalmol* **40**(5), 563--572.
- Cronly-Dillon, J.; Persaud, K. & Gregory, R. P. (1999), 'The perception of visual images encoded in musical form: a study in cross-modality information transfer.', *Proc Biol Sci* **266**(1436), 2427--2433.
- Cronly-Dillon, J.; Persaud, K. C. & Blore, R. (2000), 'Blind subjects construct conscious mental images of visual scenes encoded in musical form', *Proceedings of the Royal Society B: Biological Sciences* **267**, 2231-2238.

- Dagnelie, G.; Keane, P.; Narla, V.; Yang, L.; Weiland, J. & Humayun, M. (2007), 'Real and virtual mobility performance in simulated prosthetic vision.', *J Neural Eng* **4**(1), S92--101.
- Dobelle, W.; Mladejovsky, M.; Evans, J.; Roberts, T. & Girvin, J. (1976), '“Braille” reading by a blind volunteer by visual cortex stimulation', *Nature* **259**, 111-112.
- Dobelle, W. H. (2000), 'Artificial vision for the blind by connecting a television camera to the visual cortex.', *ASAIO J* **46**(1), 3--9.
- Dowling, J. (2005), 'Artificial human vision', *Expert Rev Med Devices* **2**(1), 73--85.
- Eckmiller, R. (1997), 'Learning retina implants with epiretinal contacts', *Ophthalmic research* **29**(5), 281.
- Eckmiller, R.; Neumann, D. & Baruth, O. (2005), 'Tunable retina encoders for retina implants: why and how', *Journal of Neural Engineering* **2**(1), S91--S104.
- Fernández, E.; Pelayo, F.; Romero, S.; Bongard, M.; Marin, C.; Alfaro, A. & Merabet, L. (2005), 'Development of a cortical visual neuroprosthesis for the blind: the relevance of neuroplasticity.', *J Neural Eng* **2**(4), R1--12.
- Fish, R. M. (1976), 'An audio display for the blind.', *IEEE Trans Biomed Eng* **23**(2), 144--154.
- Gapenne, O. & Gaussier, P. (2005), 'Suppléance perceptive et interface : une introduction', *Arob@se* **1**, 1-7.
- Gibson, J. (1986), *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates.
- Guarniero, G. (1977), 'Tactile Vision: A Personal View', *Journal of Visual Impairment and Blindness* **71**(3), 125--30.
- Gupta, S. & Petersen, K. (2002), 'Could bionic eye end blindness?', *CNN*
- Hallum, L. E.; Cloherty, S. L. & Lovell, N. H. (2008), 'Image analysis for microelectronic retinal prosthesis.', *IEEE Trans Biomed Eng* **55**(1), 344--346.
- Hallum, L. E.; Suaning, G. J.; Taubman, D. S. & Lovell, N. H. (2005), 'Simulated prosthetic visual fixation, saccade, and smooth pursuit.', *Vision Res* **45**(6), 775--788.
- Hayes, J. S.; Yin, V. T.; Piyathaisere, D.; Weiland, J. D.; Humayun, M. S. & Dagnelie, G. (2003), 'Visually guided performance of simple tasks using simulated prosthetic vision.', *Artif Organs* **27**(11), 1016--1028.
- Humayun, M. S.; de Juan, E.; Dagnelie, G.; Greenberg, R. J.; Propst, R. H. & Phillips, D. H. (1996), 'Visual perception elicited by electrical stimulation of retina in blind humans.', *Arch Ophthalmol* **114**(1), 40--46.
- Humayun, M. S.; Weiland, J. D.; Fujii, G. Y.; Greenberg, R.; Williamson, R.; Mech, J. L. B.; Cimarusti, V.; Boemel, G. V.; Dagnelie, G. & de Juan Jr, E. (2003), 'Visual perception in a blind subject with a chronic microelectronic retinal prosthesis', *Vision Research* **43**, 2573B--2581.
- Jansson, G. (1983), 'Tactile guidance of movement', *International journal of neuroscience* **19**(1), 37--46.

- Kaczmarek, K. & Haase, S. (2003), 'Pattern identification as a function of stimulation on a fingertip-scanned electrotactile display', *Neural Systems and Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Rehabilitation Engineering]* **11**(3), 269--275.
- Kaczmarek, K. & Bach-y-Rita, P. (1995), *Advanced Interface Design and Virtual Environments*, Oxford University Press, chapter Tactile Displays, pp. 349-414.
- Kajimoto; Kanno & Tachi (2006), A Vision Substitution System using Forehead Electrical Stimulation, in 'Conf. On Computer Graphics and Interactive Techniques (SIGGRAPH2006)'.
- Kanda, H.; Morimoto, T.; Fujikado, T.; Tano, Y.; Fukuda, Y. & Sawai, H. (2004), 'Electrophysiological studies of the feasibility of suprachoroidal-transretinal stimulation for artificial vision in normal and RCS rats', *Investigative ophthalmology & visual science* **45**(2), 560--566.
- Kotler, S. (2002), 'Vision Quest', *Wired*.
- Laloyaux, C. & Schmitt, C. (2003), 'Which image processing algorithms best describe the minimal amount of visual information required for image recognition?' 25th Annual Meeting of the Cognitive Science Society'.
- Lenay, C. (2002), 'Ignorance et suppléance, la question de l'espace'. Manuscrit d'habilitation à diriger des recherches.
- Lenay, C.; Gapenne, O.; Hannequin, S.; Marque, C. & Genouille, C. (2000), *La substitution sensorielle: Limites et perspectives*, chapter Toucher pour connaître. Psychologie cognitive de la perception tactile manuelle, pp. 287--306.
- Liu, W.; Fink, W.; Tarbell, M. & Sivaprakasam, M. (2005), 'Image processing and interface for retinal visual prostheses', *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, 2927--2930.
- Loomis, J. (1974), 'Tactile letter recognition under different modes of stimulus presentation', *Perception & Psychophysics* **16**(2), 401--408.
- Marr, D. (1982), *Vision: A computational investigation into the human representation and processing of visual information*, Henry Holt and Co., Inc. New York, NY, USA.
- Meijer, P. B. (1992), 'An experimental system for auditory image representations.', *IEEE Trans Biomed Eng* **39**(2), 112--121.
- Merabet, L.; Rizzo, J.; Amedi, A.; Somers, D. & Pascual-Leone, A. (2005), 'What blindness can tell us about seeing again: merging neuroplasticity and neuroprostheses', *Nature Reviews Neuroscience* **6**(1), 71--77.
- Merabet, L. B.; Battelli, L.; Obretenova, S.; Maguire, S.; Meijer, P. & Pascual-Leone, A. (2009), 'Functional recruitment of visual cortex for sound encoded object identification in the blind.', *Neuroreport* **20**(2), 132—138.
- Mossio M. et Taraborelli D. (2008) Action-dependent perceptual invariants: from ecological to sensorimotor approaches. *Consciousness and Cognition* 17, p 1324-1340.
- News, C. (), 'The Jens Naumann Story', *CBC News* <http://www.cbc.ca/sunday/sight/story3.html>.

- O'Regan, J. & Noë, A. (2002), 'A sensorimotor account of vision and visual consciousness', *Behavioral and Brain Sciences* **24**(05), 939--973.
- Perez-Fornos, A.; Sommerhalder, J.; B, R.; Safran, A. & Pelizzone, M. (2005), 'Simulation of Artificial Vision, III: Do the Spatial or Temporal Characteristics of Stimulus Pixelization Really Matter?', *Investigative Ophthalmology & Visual Science* **46**(10), 3906--3912.
- Peterman, M.; Bloom, D.; Lee, C.; Bent, S.; Marmor, M.; Blumenkranz, M. & Fishman, H. (2003), 'Localized neurotransmitter release for use in a prototype retinal interface', *Investigative ophthalmology & visual science* **44**(7), 3144--3149.
- Poirier, C.; De Volder, A.; Tranduy, D. & Scheiber, C. (2005), 'Neural changes in the ventral and dorsal visual streams during pattern recognition learning', *Neurobiol Learn Mem.*
- Pollok, B.; Schnitzler, I.; Stoerig, P.; Mierdorf, T. & Schnitzler, A. (2005), 'Image-to-sound conversion: experience-induced plasticity in auditory cortex of blindfolded adults', *Exp Brain Res* **167**, 287-291.
- Proulx, M.; Stoerig, P.; Ludowig, E. & Knoll, I. (2006), 'Seeing 'where'through the ears? Localization by sensory substitution', .
- Renier, L.; Laloyaux, C.; Collignon, O.; Tranduy, D.; Vanlierde, A. & De Volder, A. (2003), 'Perception of visual illusions with a sensory substitution system', *Perception* **32**.
- Rizzo, J.; Wyatt, J.; Loewenstein, J.; Kelly, S. & Shire, D. (2003), 'Methods and perceptual thresholds for short-term electrical stimulation of human retina with microelectrode arrays', *Investigative ophthalmology & visual science* **44**(12), 5355--5361.
- Sampaio, E.; Maris, S. & Bach-y-Rita, P. (2001), 'Brain plasticity: 'visual' acuity of blind persons via the tongue', *Brain Res* **908**(2), 204--7.
- Schmidt, E. M.; Bak, M. J.; Hambrecht, F. T.; Kufta, C. V.; O'Rourke, D. K. & Vallabhanath, P. (1996), 'Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex.', *Brain* **119** ( Pt 2), 507--522.
- Segond, H.; Weiss, D. & Sampaio, E. (2005), 'Human spatial navigation via a visuo-tactile sensory substitution system', *Perception* **34**(10), 1231.
- Smets, G. & Overbeeke, K. (1995), 'Trade-off between resolution and interactivity in spatial taskperformance', *Computer Graphics and Applications, IEEE* **15**(5), 46--51.
- Sommerhalder, J.; Rappaz, B.; de Haller, R.; Fornos, A. P.; Safran, A. B. & Pelizzone, M. (2004), 'Simulation of artificial vision: II. Eccentric reading of full-page text and the learning of this task', *Vision Research* **44**(14), 1693 – 1706.
- Taraborelli D. et Mossio M. (2008) On the relation between the enactive and the sensorimotor approach to perception. *Consciousness and Cognition*. 17(4) p 1343-1344.
- Thompson, R. W.; Barnett, G. D.; Humayun, M. S. & Dagnelie, G. (2003), 'Facial recognition using simulated prosthetic pixelized vision.', *Invest Ophthalmol Vis Sci* **44**(11), 5035--5042.
- Troyk, P.; Bak, M.; Berg, J.; Bradley, D.; Cogan, S.; Erickson, R.; Kufta, C.; McCreery, D.; Schmidt, E. & Towle, V. (2003), 'A model for intracortical visual prosthesis research', *Artificial organs* **27**(11), 1005--1015.



- Troyk, P.; Bradley, D.; Bak, M.; Cogan, S.; Erickson, R.; Hu, Z.; Kufta, C.; McCreery, D.; Schmidt, E.; Sung, S. & others (2006), 'Intracortical Visual Prosthesis Research-Approach and Progress', , 7376--7379.
- Varela, F. (1989), *Introduction aux sciences cognitives*.
- Varela, F.; Thompson, E. & Rosch, E. (1993), *L'inscription corporelle de l'esprit*, Vol. 235.
- Veraart, C.; Duret, F.; Brelén, M.; Oozeer, M. & Delbeke, J. (2004), 'Vision rehabilitation in the case of blindness.', *Expert Rev Med Devices* **1**(1), 139--153.
- Veraart, C.; Wanet-Defalque, M.; Gérard, B.; Vanlierde, A. & Delbeke, J. (2003), 'Pattern recognition with the optic nerve visual prosthesis', *Artificial Organs* **27**(11), 996-1004.
- Vurro, M.; Baselli, G.; Orabona, F. & Sandini, G. (2006), 'Simulation and Assessment of Bioinspired Visual Processing System for Epi-retinal Prostheses', , 3278--3281.
- Yagi, T.; Watanabe, M.; Ohnishi, Y.; Okuma, S. & Mukai, T. (2005), 'Biohybrid Retinal Implant: Research and Development Update in 2005', , 248--251.
- Yanai, D.; Weiland, J. D.; Mahadevappa, M.; Greenberg, R. J.; Fine, I. & Humayun, M. S. (2007), 'Visual performance using a retinal prosthesis in three subjects with retinitis pigmentosa.', *Am J Ophthalmol* **143**(5), 820--827.
- Zaghloul, K. A. & Boahen, K. (2006), 'A silicon retina that reproduces signals in the optic nerve', *J. Neural Eng.* **3**, 257B“267.
- Zrenner, E. (2002), 'The subretinal implant: can microphotodiode arrays replace degenerated retinal photoreceptors to restore vision?', *Ophthalmologica* **216 Suppl 1**, 8--20; discussion 52-3.

## Chapitre 2 : Simulation de la chaîne de traitement rétino-corticale du primate

L'objectif de ce chapitre est double. D'une part, il vise à présenter le modèle dont est issue la chaîne de traitement que nous nous proposons d'implémenter dans un dispositif de substitution. D'autre part, il vise à permettre au lecteur de comprendre en quoi ce traitement pourrait s'avérer utile dans un dispositif de suppléance visuelle.

Mon travail dans ce cadre a essentiellement été un travail de collecte et de mise en relation des différents éléments du modèle. Ce travail a abouti à la réalisation d'une simulation logicielle de la chaîne de traitement rétino-corticale. Elle a permis en particulier de vérifier la faisabilité d'une telle simulation en temps réel, et de faire en sorte que cette simulation soit paramétrable et puisse être ainsi facilement adaptée notamment à la substitution sensorielle.

Le modèle présenté ici est issue des travaux initiés au laboratoire par Jeanny Hérault dans les années 90. La première thèse à ce sujet, intitulée "Le traitement neuronal de l'information dans la rétine des vertébrés : un creuset d'idées pour la vision artificielle" a été soutenue par William Beaudot en 1994 (Beaudot, 1994). Ces travaux partent d'une analogie formelle entre le comportement des neurones de la rétine et des circuits électroniques équivalents. L'analyse des fonctions de transfert de ces circuits équivalents permet de retrouver *a posteriori* de nombreuses propriétés physiologiques du système visuel, notamment le comportement spatio-temporel des différents types de cellules ganglionnaires dans le nerf optique, et l'existence de plusieurs "voies" de transfert différant par leur comportement spatio-temporel : les voies Magnocellulaires et Parvocellulaires. Nous présenterons ces travaux dans la partie "Comportement spatio-temporel de la rétine : genèse des voies Parvo et Magnocellulaires".

Ce modèle a ensuite été complété, notamment pour rendre compte des capacités d'adaptations de la rétine qui lui permettent de fonctionner dans une large gamme de variations d'intensité lumineuse. Encore une fois, les travaux initiaux à ce sujet ont été réalisés par William Beaudot, qui propose dans sa thèse un modèle de l'adaptation des cônes à la luminance locale. Nous étendrons ce modèle à une adaptation de la réponse de la rétine aux contrastes locaux. Nous traiterons de ces propriétés dans la partie "Adaptations locales dans la rétine".

La dernière partie de ce chapitre, intitulée "Echantillonnage spatialement-variant et projections corticales" sera consacrée à la question de l'échantillonnage spatialement variant effectué par les cônes au niveau de la rétine, en lien avec le modèle de Schwartz (1977) concernant la projection des fibres visuelles au niveau du cortex visuel primaire, de type "Log-Polaire".

Avant de commencer le travail de modélisation à proprement parler, je vais présenter succinctement les grandes lignes anatomiques de la chaîne rétino-corticale du primate.

### 2.1 Anatomie générale de la chaîne rétino-corticale du primate

Le système visuel est sans doute le système sensoriel le plus développé chez le primate. A lui seul, il implique presque le quart du système nerveux central. Il est en général présenté comme une série de centres de traitement reliés entre eux par des "voies" conduisant l'information visuelle d'un centre à un autre.

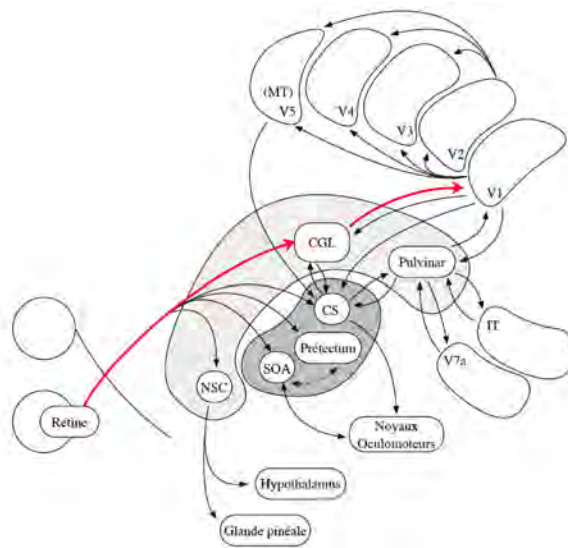


Figure 2.1 : les différentes voies du système visuel du primate et leurs interconnexions. En gras, la chaîne rétinocorticale (rétine-CGL-V1). Modifié, d'après Beaudot, 1994.

Le point d'entrée du système visuel est l'œil. Grâce au dispositif optique formé par la cornée, la pupille et le cristallin, l'œil focalise la lumière incidente au niveau de sa paroi interne.

La paroi interne de l'œil est tapissée d'une structure cellulaire appelée rétine. On considère en général que son rôle est de convertir l'énergie lumineuse en un influx nerveux. Cet influx nerveux est propagé via le nerf optique dans différentes voies (fig. 2.1).

On considère qu'environ 90% des fibres du nerf optique empruntent la voie "rétino-géniculostriée", appelée encore "voie visuelle primaire". Cette voie est aujourd'hui considérée comme la voie donnant accès à une perception qualifiée de "consciente", ou de "phénoménologique", par opposition aux autres voies manifestement dédiées au contrôle de comportements automatiques, comme le réflexe pupillaire et la régulation des rythmes circadiens (voie "photique"), ou encore les mouvements oculaires réflexes (voie "rétino-tectale" et "système optique accessoire").

On peut séparer la voie visuelle primaire en deux grandes étapes :

1. L'acheminement et le formatage de l'information visuelle depuis la rétine jusqu'au cortex visuel primaire (aire V1). Cette étape constitue ce que nous appelons la "chaîne rétinocorticale" (figure 2.2). C'est sur cette partie du système visuel que porte notre modélisation. Elle comprend l'œil, la rétine, le corps genouillé latéral, et les radiations optiques vers le cortex visuel primaire.
2. Le traitement des informations visuelles par les aires corticales et sous-corticales qui forment respectivement la voie "dorsale" et la voie "ventrale" identifiées par Ungerleider et Mishkin (1982) et dont les propriétés fonctionnelles seraient différentes, la première étant dédiée à la perception de l'espace et de l'action, et la seconde à la reconnaissance des objets (figure 2.3).

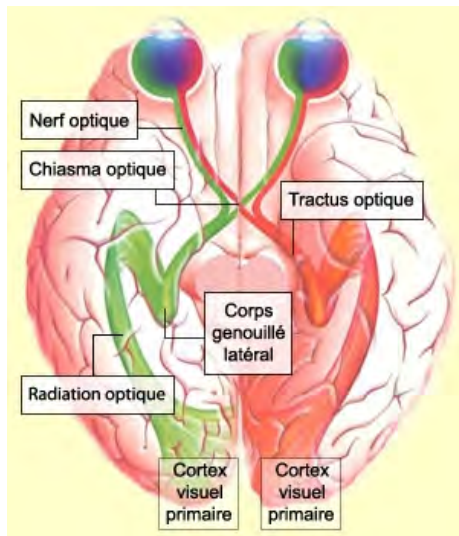


Figure 2.2 : Schéma de la chaîne rétino-corticale chez l'humain

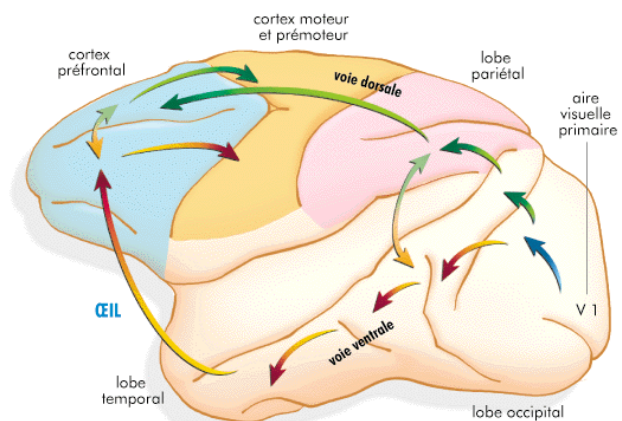


Figure 2.3 : La partie corticale et sous corticale de la voie visuelle primaire. D'après Boussaoud, 1998

Nous allons voir maintenant plus en détail les éléments constituant la chaîne rétino-corticale. Les informations présentées ici sont principalement issues de deux ouvrages de référence vers lequel le lecteur intéressé pourra se tourner ("Vision", Buser et Imbert, 1987 ; "La Vision", Rodieck, 2003), ainsi que des thèses de William Beaudot (1994) et d'Adrien Wohrer (2008) et du site Internet de référence "Webvision" (Kolb *et al.*, 2009). Le lecteur intéressé par les aspects controversés de l'anatomie rétinienne pourra par ailleurs se référer à l'article de revue de Field & Chichilnisky (2001). On trouvera également un état de l'art concernant la modélisation du système visuel primaire dans Carandini *et al.* (2005).

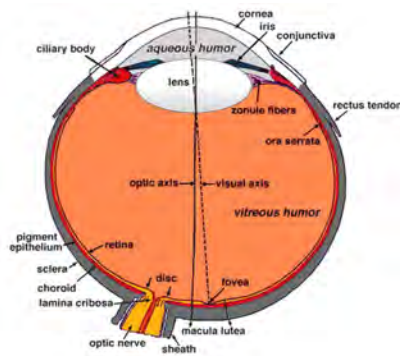


Figure 2.4: Schéma de l'œil humain

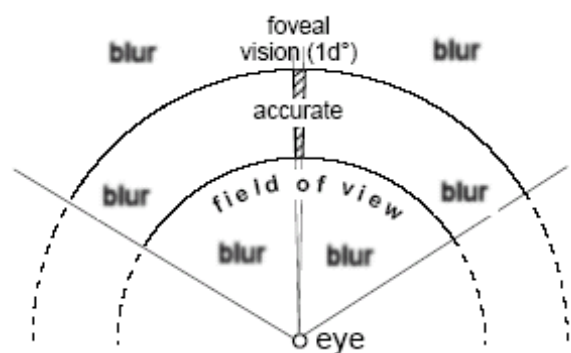


Figure 2.5: Champ visuel humain

## 2.1.1 L'œil

L'œil (fig. 2.4) est le premier élément de la chaîne, il reçoit les rayons lumineux réfléchis par les objets et les focalise sur la rétine. Chaque œil est maintenu en place par trois paires de muscles extra-oculaires : une paire de muscles obliques et deux paires de muscles droits qui permettent les mouvements de saccades oculaires.

La pupille est l'ouverture centrale de l'iris. Cette membrane colorée présente un muscle de disposition circulaire qui modifie le diamètre de la pupille en ajustant la quantité de lumière qui entre dans l'œil.

La cornée est la partie antérieure et transparente de l'œil qui couvre l'iris et la pupille. Elle est la première lentille du globe oculaire et assure la part la plus importante de la focalisation. L'autre lentille est le cristallin, situé derrière l'iris. Sa forme varie selon la contraction ou la relaxation des ligaments par l'action des muscles ciliaires, ce qui permet la focalisation des images selon la distance. Signalons ici que l'image rétinienne n'est nette qu'au niveau de la fovéa, la zone centrale de la rétine couvrant environ 1° de champ visuel, le champ visuel humain total approchant les 140° (fig. 2.5).

## 2.1.2 La rétine

La rétine est un feuillet de tissu neural épais d'environ 0.4 millimètre qui tapisse l'intérieur de la face postérieure du globe oculaire et s'étend en avant jusqu'à "l'ora serrata" (fig. 2.6).

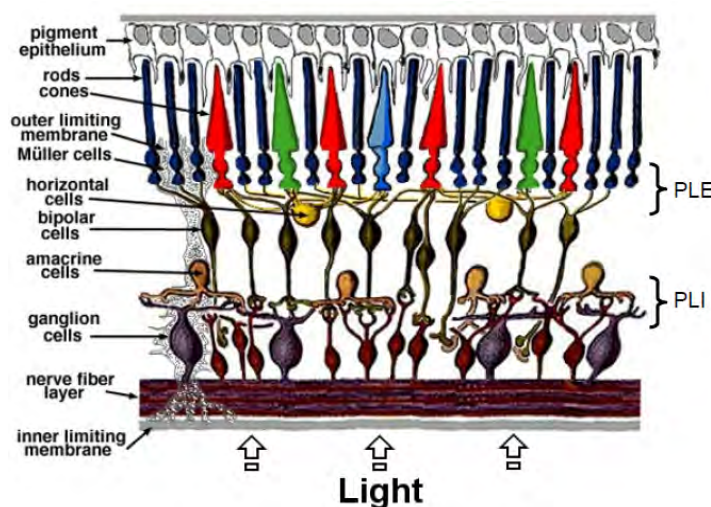


Figure 2.6 : schéma de la rétine humaine. Modifié, d'après webvision

La rétine comprend principalement cinq classes de neurones organisés en couches successives : les photorécepteurs, les cellules horizontales, les cellules bipolaires, les cellules amacriennes et enfin les cellules ganglionnaires qui forment le nerf optique. L'image formée par le système optique de l'œil sur la rétine est convertie en potentiel électrique par les photorécepteurs. Ce potentiel électrique est relayé par les cellules bipolaires vers les cellules ganglionnaires qui forment la sortie de la rétine : le nerf optique.

Les cellules horizontales et amacriennes interagissent avec les autres cellules de la rétine et influent sur la propagation du signal visuel. Ces interactions se font à deux niveaux dans la rétine :

1- La couche plexiforme externe (PLE), où les photorécepteurs et les cellules horizontales entrent en contact avec les cellules bipolaires.

2 – La couche plexiforme interne (PLI), où les cellules bipolaires entrent en contact avec les cellules amacriennes pour interagir avec les cellules ganglionnaires.

### Les photorécepteurs :

Les cellules photosensibles de la rétine sont les photorécepteurs. Contrairement à ce que l'on pourrait croire, les photorécepteurs ne sont pas à la surface extérieure de la rétine en contact direct avec la lumière, mais au contraire au niveau de sa surface interne.

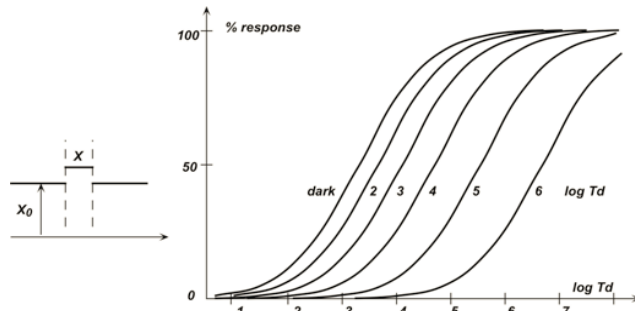


Figure 2.7 : réponse des cônes en fonction de l'éclairement de l'arrière plan. D'après Valeton et van Norren, 1983.

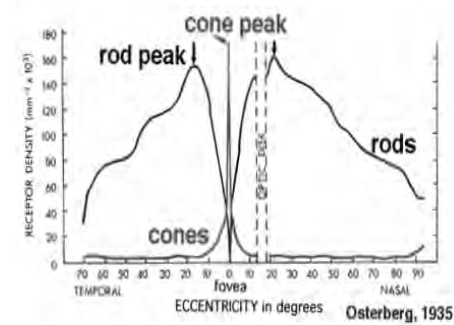


Figure 2.8: densité des cônes et des bâtonnets dans la rétine humaine. D'après Osterberg, 1935.

Il existe deux grandes catégories de photorécepteurs : les cônes et les bâtonnets. Les bâtonnets sont capables de transmettre des variations d'intensité lumineuse dans des conditions d'obscurité, mais perdent cette propriété en pleine lumière. Ils sont responsables de la vision nocturne. Les cônes sont légèrement moins sensibles que les bâtonnets mais sont adaptés pour travailler à la pleine lumière d'un jour ensoleillé. Ils sont responsables de la vision diurne. Par souci de simplicité, nous ne traiterons dans la suite que de la circuiterie impliquant les cônes.

L'existence de deux types de photorécepteurs ne suffit pas à expliquer comment ce système de capteur réussi à travailler sur toute la gamme des luminosités utiles. En vision scotopique en particulier, la gamme de luminosité varie sur 6 ordres de grandeurs, alors que les cônes ne peuvent en coder que 2. Pour palier à cela, la dynamique de réponse des cônes est elle-même modulée en fonction de la luminosité locale : c'est ce que l'on nomme la compression adaptative. La figure 2.7 montre la courbe de réponse des photorécepteurs pour différentes valeurs de luminosité ambiante ( $X_0$ ). La fonction de transduction est à peu près logarithmique, et la courbe de réponse du cône se translate sans changer de forme. Nous reviendrons sur ces aspects dans la partie sur l'adaptation dans la rétine.

Autre point anatomique remarquable : la densité des cônes et des bâtonnets n'est pas constante sur la rétine (fig. 2.8). En particulier, la rétine présente une zone, nommée fovéa, où la densité des cônes est maximale et dans laquelle on ne trouve pas de bâtonnets. Nous reviendrons en détail sur cette question dans la dernière partie de ce chapitre.

On identifie par ailleurs trois types de cônes selon leur sensibilité à la longueur d'onde de la lumière incidente. Les cônes "S" (pour "short wavelength") sont sensibles aux courtes longueurs d'ondes (bleues), les cônes "M" aux longueurs d'ondes moyennes (vertes), et les cônes "L" aux grandes longueurs d'ondes (rouge). Ces différences sont responsables de la capacité de l'humain à discriminer des couleurs.

Enfin, on notera que deux cônes voisins sont couplés électriquement par une synapse appelée jonction GAP.

### Les cellules horizontales :

La particularité des cellules horizontales est de posséder une arborisation dendritique large. Elles interagissent avec plusieurs photorécepteurs, jusqu'à une vingtaine dans les zones périphériques de la rétine. Elles sont également liées entre elles par des jonctions GAP. Elles génèrent de ce fait des interactions latérales étendues dans la PLE, ce qui leur vaut leur nom de cellules "horizontales".



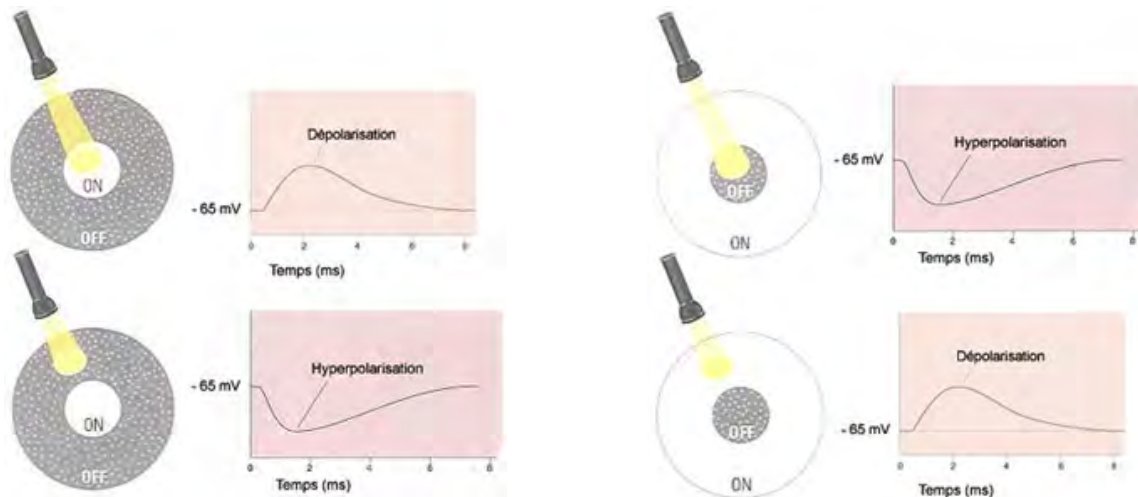


Figure 2.9 : champs récepteurs des cellules bipolaires. A gauche, cellule bipolaire ON, à droite cellule bipolaire OFF.

### Les cellules bipolaires :

Les cellules bipolaires transmettent le signal nerveux de la couche plexiforme externe vers la couche plexiforme interne.

On distingue essentiellement deux types de cellules bipolaires de cônes : les bipolaires "Midget" et les bipolaires "Diffuse". L'arborisation dendritique des cellules Midget est restreinte à un seul cône, du moins dans la fovéa. A l'inverse, les bipolaires Diffuse peuvent faire des contacts avec plusieurs cônes. L'arborisation dendritique des cellules bipolaires Diffuse augmente avec l'excentricité.

L'autre distinction importante entre les cellules bipolaires n'est pas d'ordre anatomique. Elle concerne leur réponse caractéristique à une stimulation lumineuse. Les cellules bipolaires présentent en effet un champ récepteur biphasique de type "Centre/Périphérie" (figure 2.9).

On distingue ainsi les cellules bipolaires de type ON de celles de type OFF. Les bipolaires ON sont activées par une stimulation au centre de leur champ récepteur et sont inhibées par une stimulation sur le pourtour. A l'inverse les cellules bipolaires OFF sont inhibées par une stimulation au centre de leur champ récepteur, et sont activées par une stimulation sur le pourtour.

### La couche plexiforme externe :

La couche plexiforme externe est le lieu d'interactions entre les photorécepteurs, les cellules horizontales et les cellules bipolaires, qui forment ce que l'on appelle une triade synaptique (fig. 2.10). On distingue deux types de synapses dans la triade : les synapses invaginant (qui s'insèrent dans un renflement du pédicule du cône) et les synapses non-invaginant, qui contactent la surface du pédicule.

Le signal des bipolaires ON est formé au niveau des synapses invaginant. Elles sont activées par les photorécepteurs et inhibées par les cellules horizontales. Le signal des bipolaires OFF est formé au niveau des synapses non-invaginant. A l'inverse, elles sont inhibées par les photorécepteurs et activées par les cellules horizontales. Ces différences sont à l'origine des différences de polarités observées dans les champs récepteurs des cellules bipolaires ON et OFF.

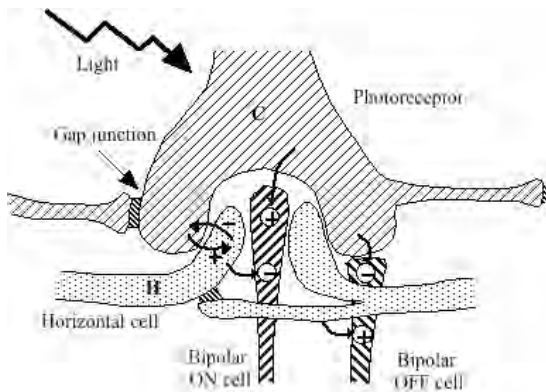


Figure 2.10 : la triade synaptique, dans la couche plexiforme externe. D'après Hérault, sous presse.

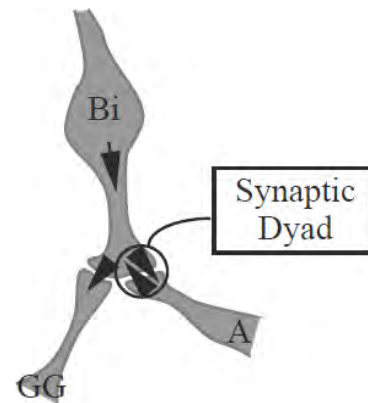


Figure 2.11 : schéma d'une interconnections fréquente dans la couche plexiforme interne impliquant une dyade synaptique. D'après Beaudot, 1994

### Les cellules amacrines :

A la manière des cellules horizontales dans la couche plexiforme externe, les cellules amacrines interagissent avec les cellules bipolaires et les cellules ganglionnaires dans la couche plexiforme interne. Le rôle exact des cellules amacrines est néanmoins bien moins connu que celui des cellules horizontales. Leur étude est compliquée par le grand nombre de type différents d'amacrines : on estime qu'il en existe de 20 à 40 sortes.

Globalement on attribue aux cellules amacrines un rôle de régulation de l'activité des cellules ganglionnaires. Elles pourraient également intervenir dans la régularisation du contraste spatial et temporel.

### Les cellules ganglionnaires :

Les cellules ganglionnaires forment la dernière couche de cellules dans la rétine. Leurs axones se regroupent au niveau de la papille pour former le nerf optique.

Comme les cellules bipolaires, on identifie principalement deux types de cellules ganglionnaires : les cellules ganglionnaires Midget et les cellules ganglionnaires Parasol. En fovéa, les cellules ganglionnaires Midget forment une connexion "1 pour 1" avec les cellules bipolaires de même type. Le circuit Midget est donc direct dans le centre de la rétine : une ganglionnaire Midget contacte une seule cellule bipolaire Midget, qui elle-même ne contacte qu'un cône. A l'inverse, les cellules ganglionnaires Parasol possèdent une arborisation dendritique large, et de plus en plus large à mesure que l'on s'éloigne de la fovéa. Par ailleurs, elles contactent également des cellules bipolaires Diffuse.

Tout comme les cellules bipolaires, les cellules ganglionnaires présentent également un champ récepteur biphasique de type "Centre/Contour", et de polarité ON ou OFF. Les cellules ganglionnaires de type ON et OFF contactent respectivement des cellules bipolaires de même polarité. Par ailleurs, la taille globale des champs récepteurs des cellules Diffuse est plus grand que celui des cellules Midget.

Plus récemment, on a identifié un troisième type de cellule ganglionnaire dite "Bistratifiée". Cette dernière contacte à la fois des cellules bipolaires de polarité ON et de polarité OFF. Globalement, ce type de cellule répond à une variation de la stimulation, que ce soit au centre ou en périphérie.



Il existe enfin une dernière différence entre les cellules ganglionnaires Midget et Parasol concernant leur comportement temporel. Les cellules ganglionnaires Parasol ont une réponse transitoire, c'est à dire qu'elles ne répondent qu'à la variation de stimulation. A l'inverse, les ganglionnaires Midget ont un comportement tonique, c'est à dire dont le taux de décharge est maintenu tant que la stimulation est maintenue.

Nous verrons dans la suite du propos que les cellules ganglionnaires peuvent aussi être distinguées en fonction de leur destination dans le corps genouillé latéral.

### **La couche plexiforme interne :**

La couche plexiforme interne est le siège des interactions entre les cellules bipolaires, les amacrines et les cellules ganglionnaires. Vu le nombre de cellules amacrines répertoriées, les modes d'interaction entre ces cellules ne sont pas tous connus. On identifie néanmoins une structure récurrente appelée dyade synaptique (fig. 2.11).

On considère en général que les cellules amacrines ont un rôle de régulation de l'activité des cellules ganglionnaires, et en particulier, qu'elles accentuent le caractère transitoire des cellules ganglionnaires Parasol.

## **2.1.3 Le corps genouillé latéral**

Les nerfs optiques de l'œil gauche et de l'œil droit se rejoignent au niveau du chiasma optique. Là, les fibres visuelles sont aiguillées en fonction de la partie du champ visuel qu'elles transmettent. Les fibres transmettant l'information concernant l'hémichamp visuel gauche sont aiguillées vers l'hémisphère droit du cerveau, celles concernant l'hémichamp visuel droit sont aiguillées vers l'hémisphère gauche. On notera néanmoins que les fibres en provenance de l'œil droit ne sont pas mélangées avec celles en provenance de l'œil gauche. Cette séparation subsiste au moins jusque dans l'aire visuelle primaire.

Le tractus optique transmet l'influx nerveux du chiasma optique jusqu'au corps genouillé latéral (CGL). Le CGL est une structure thalamique relais. Chez le primate, le CGL comporte 6 couches concentriques à convexité dorsale (fig. 2.12). Chaque couche mesure environ un demi millimètre d'épaisseur et reçoit l'information de l'hémichamp rétinien d'un seul œil.

Les projections des cellules ganglionnaires sur les couches du CGL respectent par ailleurs la topologie de la rétine de sorte que des fibres issues de cellules ganglionnaires voisines dans la rétine contactent des cellules voisines dans le CGL. De plus, les cartes rétinotopiques des différentes couches sont superposées en correspondance exacte de sorte qu'un point particulier dans le champ visuel se projette sur une ligne de projection à l'intérieur du CGL traversant toutes les couches (représenté sur la figure 2.12).

Voyons maintenant le détail de ces couches. Les deux premières couches (1-2) sont constituées de cellules relativement grosses : on les nomme pour cela couches Magnocellulaires (du latin "Magnus" : grand). Elles reçoivent principalement des projections des cellules ganglionnaires de type Parasol. Les quatre couches supérieures (3-6) sont constituées de corps cellulaires de plus petite taille : on les nomme pour cela couches Parvocellulaires (du latin "Parvus" : petit). Elles reçoivent principalement des projections des cellules ganglionnaires de type Midget. Depuis peu, on identifie dans le CGL un troisième type de couches constituées de corps cellulaires de très petite taille : ce sont les couches dites Koniocellulaire (du grec Konis = poussière). Elles sont situées entre les couches parvocellulaires et magnocellulaires. Pour Rodieck, elles constitueraient des sous-couches plutôt que des couches à part entière (figure 2.13). Ces sous-couches recevraient



Figure 2.12 : corps genouillé du primate. Rodieck 2003, d'après Hubel et Wiesel, 1977

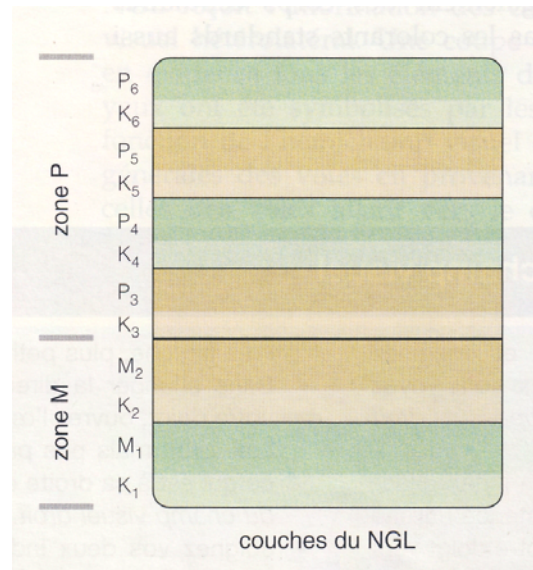


Figure 2.13 : schéma de l'organisation en couches du CGL, d'après Rodieck 2003

notamment des projections des cellules ganglionnaires bistratifiées.

Cette séparation des cellules ganglionnaires en fonction de leur type a conduit à définir trois voies différentes, la voie Parvocellulaire, la voie Magnocellulaire et la voie Koniocellulaire. Ces voies sont maintenues au moins jusqu'au cortex visuel primaire et présentent des propriétés fonctionnelles distinctes. La voie Parvocellulaire (Parvo) est associée à des champs récepteurs petits et véhicule de ce fait une version détaillée, précise, de la stimulation rétinienne. Associé à cela, sa réponse temporelle relativement lente mais tonique lui confère un intérêt particulier dans les processus de reconnaissance et d'analyse précise de l'image. La voie Magnocellulaire (Magno), à l'inverse, est associée à des champs récepteurs larges et véhicule de ce fait une version grossière de la stimulation rétinienne. Cette voie est en revanche plus rapide que la voie Parvocellulaire, et répond de manière transitoire. Elle serait pour ces raisons plutôt dédiée à l'information liée au contexte et au mouvement. Les propriétés fonctionnelles de la voie Koniocellulaire (Konio) sont, elles, beaucoup moins bien connues. Elle jouerait un rôle prépondérant dans la transmission des informations associées à la couleur (cônes S).

## 2.1.4 Les radiations optiques et la projection sur le cortex visuel primaire

Les radiations optiques constituent la dernière étape de la chaîne rétino-corticale. Elles véhiculent le signal visuel depuis le corps genouillé latéral et se projettent sur le cortex visuel primaire.

La projection du champ visuel sur le cortex ne correspond pas, comme on pourrait l'imaginer, à une simple homothétie. La rétinotopie toutefois est bien conservée, c'est à dire que les projections de deux points proches dans le champ visuel seront proches au niveau du cortex. Cette projection du champ visuel sur le cortex combine deux phénomènes :

1. Une magnification de la partie centrale du champ visuel. Comme on le constate sur la figure 2.14, la première moitié du cortex visuel correspond seulement aux 10 premiers degrés du

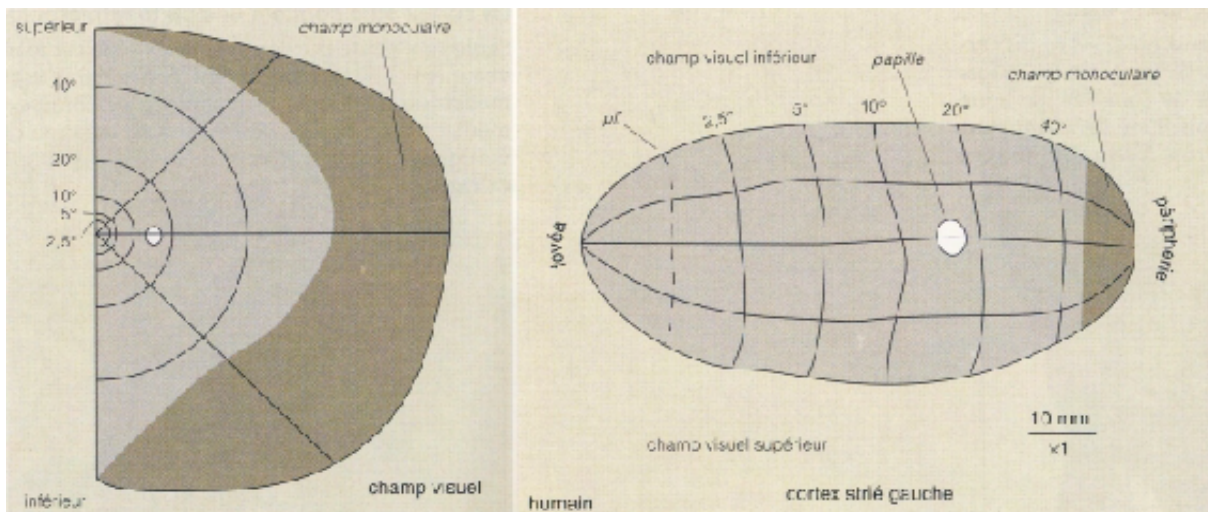


Figure 2.14 : Projection du champ visuel sur le cortex visuel primaire chez le primate. Rodieck, 2003, d'après Horton et Hoyt, 1991.

champ visuel. Ce phénomène est intimement lié à la décroissance de la densité des photorécepteurs au niveau de la rétine, ainsi qu'à l'augmentation de l'arborisation dendritique des cellules bipolaires et ganglionnaires avec l'excentricité.

2. Une transformation angulaire : les cercles concentriques représentés au niveau du champ visuel deviennent plus ou moins des droites à partir de 2.5° d'excentricité. Cette transformation est liée à la projection des radiations optiques sur le cortex visuel. Cette transformation est vraisemblablement liée à la projection des radiations optiques sur le cortex.

## 2.2 Comportement spatio-temporel de la rétine, genèse des voies "Magno" et "Parvo"

Passons maintenant à la modélisation à proprement parler de la chaîne rétinocorticale, et focalisons nous dans un premier temps sur le comportement linéaire de la rétine.

Le comportement spatio-temporel est sans doute l'un des éléments les plus étudiés au niveau de la rétine. Cette dernière agit en effet comme un aiguillage et transmet le signal visuel selon diverses voies en fonction de son contenu spatio-temporel. Comme nous l'avons vu, on identifie dès le nerf optique deux voies principales : la voie Magnocellulaire et la voie Parvocellulaire. La voie Koniocellulaire, identifiée plus récemment, est encore mal connue.

Fonctionnellement, la voie Parvocellulaire transmet les hautes fréquences spatiales de l'image et présente une réponse temporelle tonique. On l'associe en général aux tâches liées à la reconnaissance des objets. À l'inverse, la voie Magnocellulaire transmet les basses fréquences spatiales et présente une réponse temporelle transitoire. On l'associe en général à la détection du mouvement et aux informations de contexte. Cette séparation des différentes composantes du signal est vraisemblablement une base du fonctionnement du système visuel. Elle fait notamment écho aux notions de voies ventrales et dorsales identifiées par Ungerleider et Mishkin (1982), respectivement dédiées à la reconnaissance et à l'action.

L'essentiel des travaux présentés ici sont issus de la thèse de William Beaudot (1994). Le principe de ce modèle consiste à identifier formellement les structures neurales de la rétine avec des

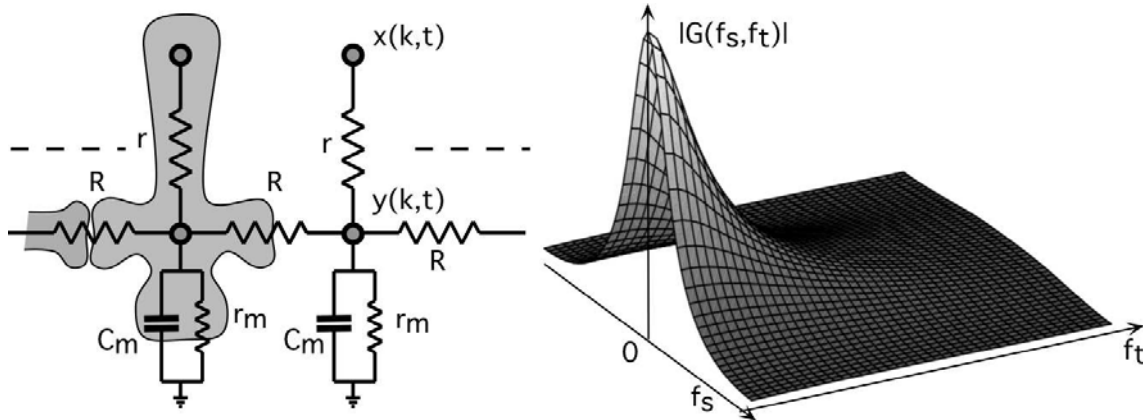


Figure 2.15 : le modèle électrique d'une couche de neurones liés entre eux par des jonctions Gap (à gauche). À droite, fonction de transfert associée à cette couche.

circuits électroniques. D'après Beaudot (1993), qui s'inspire des travaux de Mead et Mahowald (1988) un neurone peut être modélisé par un générateur de voltage présentant une résistance interne  $r$  et un circuit de fuite modélisant les propriétés membranaires de la cellule, composé d'une capacité  $C_m$  et d'une résistance  $r_m$  (fig. 2.15). Dans le cas d'un photorécepteur, le potentiel électrique est issu de la transduction de l'énergie lumineuse en énergie électrique. Dans le cas des autres cellules de la rétine, il provient des interactions synaptiques avec les cellules de la couche précédente. On calcule ensuite la fonction de transfert de ces circuits qui caractérise le comportement linéaire de la structure considérée.

Dans un premier temps, nous nous intéresserons à la couche des photorécepteurs, nous proposerons ensuite un modèle de la couche plexiforme externe, incluant les cellules horizontales et les cellules bipolaires. Nous présenterons enfin le modèle de la couche plexiforme interne et ses implications dans le comportement global de la rétine et dans la genèse des voies Magno et Parvo.

### 2.2.1 Modèle générique d'une couche de cellules rétinienne

Considérons pour débuter un réseau bidimensionnel de neurones reliés entre eux par des jonctions Gap. Une rangée de ce réseau est illustrée figure 2.15. On peut calculer, sur la base de ce modèle électronique, la fonction de transfert d'une telle couche, en fonction des fréquences spatiales ( $f_s$ ) et temporelles ( $f_t$ ). Elle s'écrit :

$$G(f_s, f_t) = \frac{1}{1 + \beta + 2\alpha(1 - \cos(2\pi f_s)) + j2\pi\tau f_t}$$

$$\text{avec : } \alpha = \frac{r}{R}, \quad \beta = \frac{r}{r_m} \quad \text{et} \quad \tau = r.C_m$$

Le module de cette fonction de transfert est représenté fig 2.15 (à droite). Il s'agit d'un filtre spatio-temporel passe-bas. Ce filtre est par ailleurs à variables non séparables, c'est à dire qu'il ne peut pas s'écrire comme le produit d'un filtre spatial par un filtre temporel.

La fonction de transfert de cette couche dépend de trois paramètres : une constante spatiale  $\alpha$ , une constante temporelle  $\tau$  et enfin un paramètre de gain  $\beta$ . Ces paramètres sont déterminés par la valeur des résistances et des capacités dans le circuit. Cette structure constitue le maillon de base de notre modèle. Nous l'utiliserons pour modéliser la couche des photorécepteurs, celle des cellules horizontales et enfin celle des cellules ganglionnaires, qui seront respectivement caractérisées par

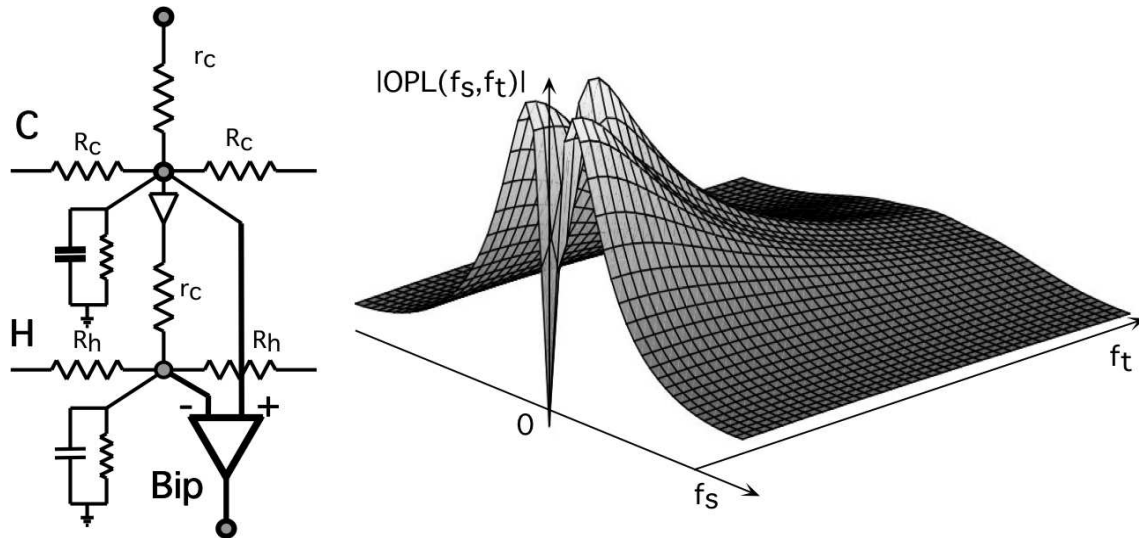


Figure 2.16 : le modèle électrique de la couche plexiforme externe (à gauche). A droite, la fonction de transfert équivalente.

un jeu particulier de constantes ( $\alpha$ ,  $\beta$ ,  $\tau$ ). Les cellules horizontales par exemple présentent, comme nous l'avons mentionné, des connexions latérales plus fortes que les photorécepteurs. La couche des cellules horizontales sera donc caractérisée par des constantes spatiales et temporelles plus fortes que celle des photorécepteurs.

Par ailleurs, pour les photorécepteurs, la différence de potentiel sera issue de la conversion de l'énergie lumineuse, qui sera assimilée dans la simulation à la valeur des niveaux de gris de l'image d'entrée. Pour les autres cellules, elle sera issue de l'interaction synaptique avec la couche précédente, c'est à dire, dans la simulation, avec les niveaux de gris de l'image de sortie de la couche précédente.

### Implémentation :

En pratique, ce filtrage est approximé par un filtre récursif causal/anticausal du premier ordre inclus dans une boucle temporelle (Beaudot, 1994), ce qui limite le coût de calcul à 5 opérations par pixels et par unité de temps, tout en préservant le caractère non séparable du comportement temporel et spatial.

Par ailleurs, on préférera au paramètre  $\alpha$  le paramètre  $k = \alpha^2$  qui correspond à la constante spatiale du filtre numérique en pixel (Beaudot, 1994). Le paramètre  $\tau$ , dans la version numérique du filtre correspond, lui, à la constante de temps exprimée en nombre d'images.

## 2.2.2 Modélisation de la couche plexiforme externe

La couche plexiforme externe est le lieu d'interactions entre les photorécepteurs, les cellules horizontales et les cellules bipolaires, au sein de la triade synaptique. Dans notre cas, nous modélisons la triade synaptique par une différence entre la sortie des cônes et celle des cellules horizontales. Le schéma électrique équivalent de ce modèle est présenté figure 2.16 ainsi que la fonction de transfert associée. Elle correspond à un filtre passe-bande spatio-temporel.

Une telle différence produit par ailleurs des valeurs négatives incompatible avec un codage neural. Nous considérons donc la sortie des cellules bipolaires ON comme la partie positive de cette



différence et la sortie des cellules bipolaires OFF comme sa partie négative.

$$\begin{cases} BipON = \text{partie positive} (Photorécepteurs - Horizontales) \\ BipOFF = \text{partie positive} (Horizontales - Photorécepteurs) \end{cases}$$

On retrouve de ce fait le schéma des connections dans la triade synaptique et la polarité des cellules bipolaires.

### Implémentation :

En pratique, nous simulons donc la PLE en calculant la différence entre l'image issue des Photorécepteurs et l'image issue des Horizontales. La partie positive et la partie négative de cette différence donnent respectivement l'image des bipolaires ON et OFF. La sortie des différentes couches de cellules interagissant dans la PLE est présentée figure 2.17.

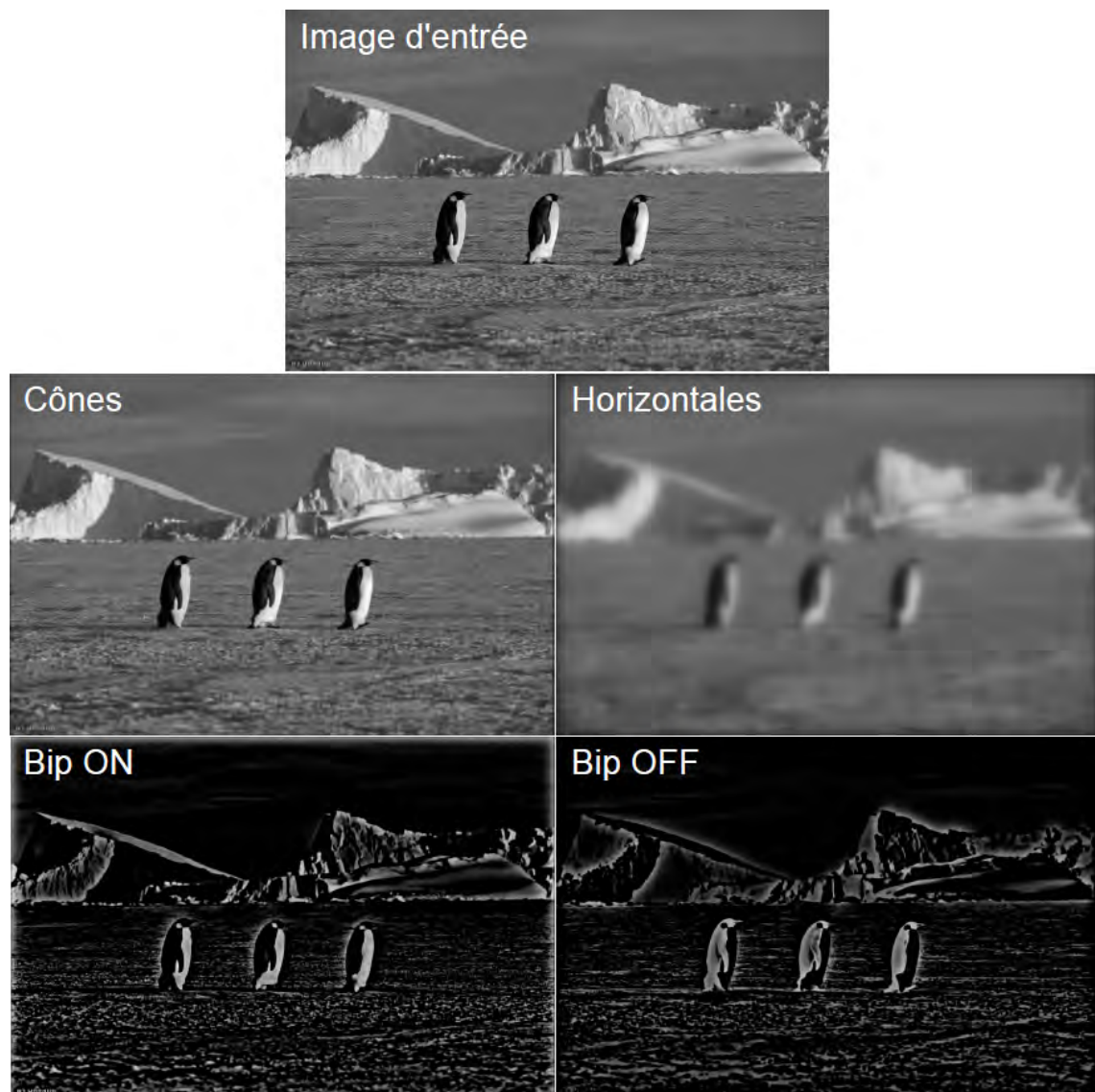


Figure 2.17: illustration de la sortie des différentes cellules interagissant dans la couche plexiforme externe. Image d'origine tirée de "La marche de l'empereur", Luc Jacquet, 2004.

## 2.2.3 Modélisation de la couche plexiforme interne

La couche plexiforme interne est le lieu d'interaction entre les cellules bipolaires, les cellules amacrines et les cellules ganglionnaires, qui forment le nerf optique. Nous allons considérer dans un premier temps la modélisation des cellules ganglionnaires Midget et Parasol, puis nous étudierons leur interaction avec les amacrines

### 2.2.3.1 Ganglionnaires Midget et Parasol : la clé de la séparation voie Magno / voie Parvo

Comme nous l'avons mentionné, les cellules ganglionnaires Midget (ON ou OFF) contactent uniquement des cellules bipolaires Midget (de même polarité), elles même liées à un seul photorécepteur. A l'inverse, les cellules ganglionnaires Parasol (ON ou OFF) contactent des cellules bipolaires Diffuse (de même polarité) qui forment des contacts avec plusieurs photorécepteurs. Les cellules ganglionnaires bistratifiées font, enfin, des connexions multiples avec les cellules bipolaires Diffuse de type ON et OFF.

Pour cela, on considère que le signal issu des ganglionnaires Midget est la copie du signal issue des cellules bipolaires (ON et OFF respectivement). On modélise en revanche les connexions multiples des cellules Parasol par un filtrage spatial passe-bas. On modélisera enfin les cellules bistratifiées en faisant la somme des cellules ganglionnaires Parasol de type ON et OFF.

Il serait difficile, à ce niveau, d'interpréter directement les différentes fonctions de transfert caractérisant le comportement des différents types de cellules ganglionnaires. A partir de cette fonction de transfert, en revanche, nous avons calculé la réponse en fréquence spatiales de chaque type de cellule, en fonction du temps (réponse indicielle).

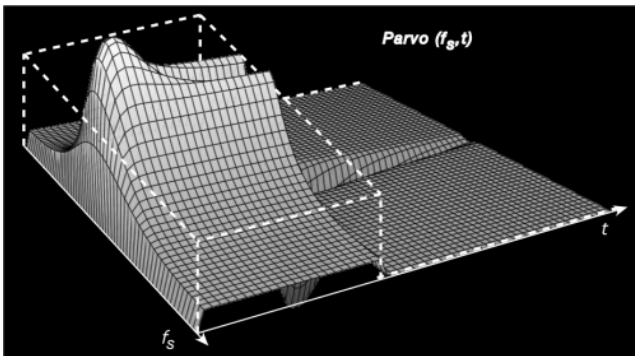


Figure 2.18 : réponse indicielle du modèle de la voie Parvo.

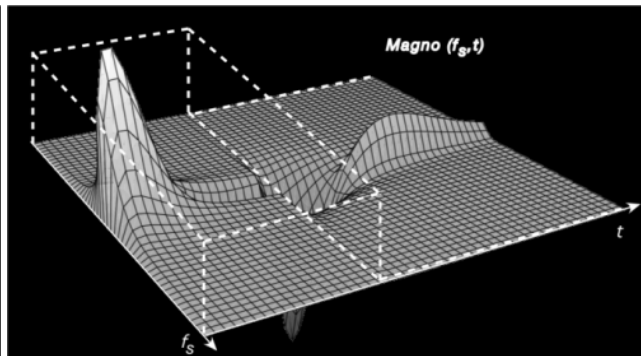


Figure 2.19 : réponse indicielle du modèle de la voie Marvo.

La réponse indicielle des cellules ganglionnaires Midget, telle que nous la modélisons ici, est présentée figure 2.18. Elle correspond d'ailleurs à la réponse indicielle des cellules bipolaires, puisque nous considérons que le signal des ganglionnaires Midget est une copie du signal des bipolaires Midget. On remarquera en premier lieu que la réponse des cellules Midget est maintenue durant toute la stimulation. Par ailleurs, les hautes fréquences de l'image sont conservées et, après un certain temps, les basses fréquences de l'image sont atténuées. On retrouve donc bien les propriétés observées au niveau de la voie Parvocellulaire.

La réponse indicielle des cellules ganglionnaires Parasol telle que nous les modélisons ici est présentée figure 2.19. Elle correspond d'ailleurs à la réponse indicielle présenté à la figure 2.18 dont

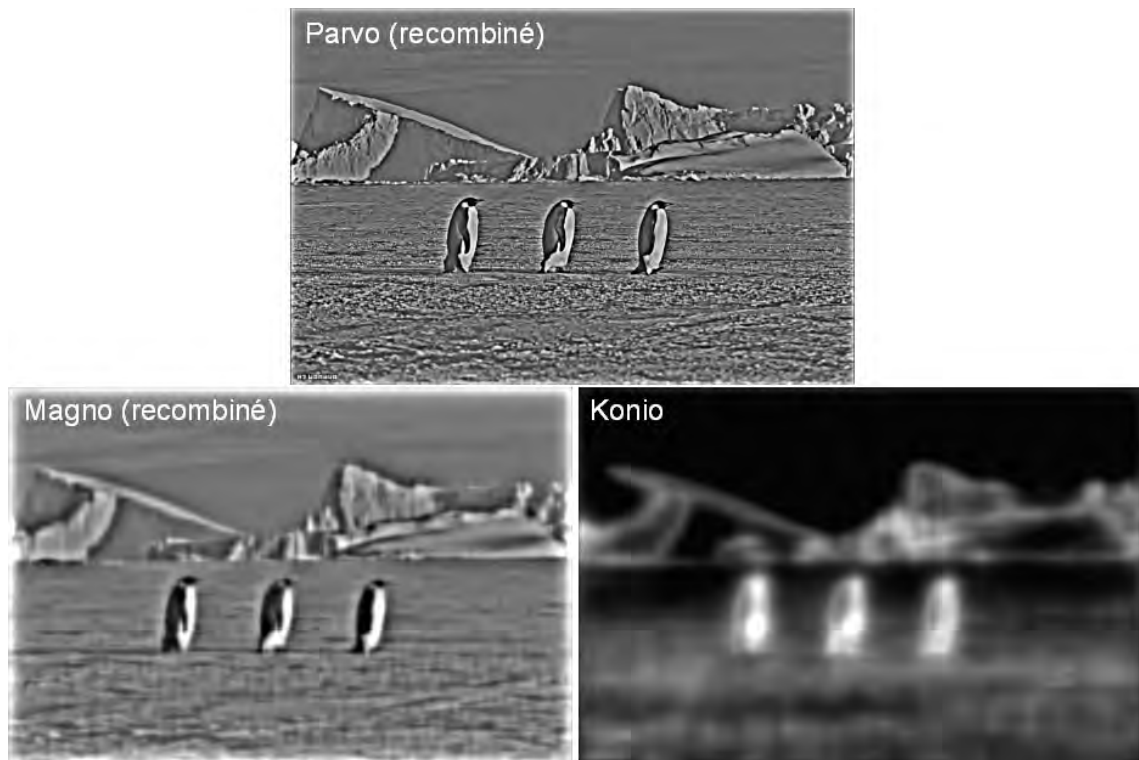


Figure 2.20 : simulation des voies Magno (ganglionnaires Parasol), Parvo (ganglionnaires Midget) et Magno ON/OFF (ganglionnaires bistratifiées).

on n'aurait gardé que les basses fréquences spatiales. On remarque notamment que la réponse des cellules Parasol est transitoire. Par ailleurs, le signal transmis par ce type de cellule ne contient que les basses fréquences du signal. On retrouve donc les propriétés observées au niveau de la voie Magnocellulaire.

Le comportement des cellules bistratifiées, qui font la somme des contributions ON et OFF des ganglionnaires parasol, n'est pas linéaire (cette somme correspond en fait à une valeur absolue sur la différence de deux signaux) et ne peut donc pas être analysée avec les mêmes outils. On notera en revanche que leur comportement transitoire dérive du comportement transitoire observé pour les cellules ganglionnaires Parasol, et que leurs champs récepteurs sont de types ON-OFF, conformément aux observations à leur égard.

#### **Remarque :**

Dans la partie consacrée à la PLE, nous n'avons pas considéré la différence entre les cellules bipolaires de type Midget et celles de type Diffuse. L'arborisation dendritique de ces dernières génère pourtant un filtrage passe-bas spatial du signal. Toutefois, s'agissant de filtrages de type linéaire, les différents opérateurs sont commutatifs. Le filtrage passe-bas spatial opéré au niveau des cellules ganglionnaires Parasol inclut donc ici le filtrage passe-bas associé aux bipolaires Diffuse.

#### **Implémentation :**

Pour des raisons pratiques, le filtrage passe-bas spatial sur le signal des bipolaires est réalisé à l'aide du filtre récursif utilisé pour les autres couches. Le paramètre temporel  $\tau$  est fixé à 0.

La sortie des différents types de cellules ganglionnaires est présentée figure 2.20. Pour mieux



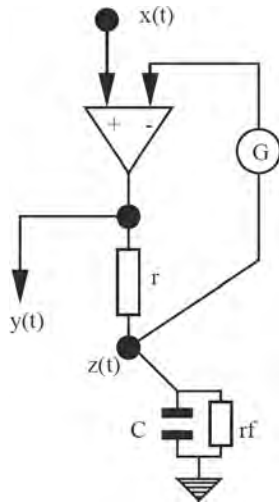


Figure 2.21 : schéma de la dyade synaptique. D'après Beaudot, 1994.

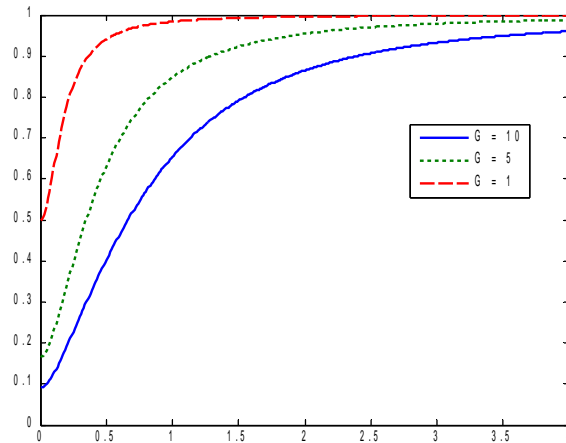


Figure 2.22 : module de la fonction de transfert de la dyade synaptique, pour différentes valeurs du gain G.

visualiser les propriétés de ces voies, nous avons combiné les voies ON et OFF dans une même image. Le niveau 0 est donc le niveau moyen (gris). Les parties plus sombres sont issues de la voie OFF, et les parties plus claires issues de la voie ON.

### 2.2.3.2 Interaction avec les cellules amacrines

Le nombre, la morphologie, le type d'interaction et la fonction des cellules amacrines est encore aujourd'hui mal connu. On tient néanmoins en général les cellules amacrines pour responsables du comportement transitoire des cellules ganglionnaires Parasol. Bien que nous ayons déjà identifié une origine possible de ce comportement transitoire, nous allons tout de même nous pencher sur cette hypothèse, pour établir si les amacrines peuvent avoir un rôle dans le renforcement de ce comportement transitoire.

On trouve un élément récurrent dans les interactions entre les cellules amacrines et les autres cellules de la PLI : la dyade synaptique (fig 2.11). La modélisation de cette dyade permet de proposer une première analyse de l'effet des cellules amacrines au niveau de la couche plexiforme externe.

La figure 2.21 précise le schéma électronique équivalent de la dyade synaptique, proposée par Beaudot.

La fonction de transfert de la dyade synaptique s'écrit alors :

$$G(f_t) = 1 - \frac{G}{1 + G + \beta + j 2 \pi \tau f_t}$$

Le module de cette fonction de transfert est représenté figure 2.22, pour différentes valeurs du gain G. Elle correspond à un filtre temporel passe-haut, compatible avec l'hypothèse selon laquelle les cellules amacrines auraient un lien avec le comportement transitoire des cellules ganglionnaires. Pour notre part, nous pensons que les cellules amacrines viennent accentuer le comportement transitoire des cellules ganglionnaires Parasol.

**Implémentation :**

Ce filtre est formellement identifiable au filtre de base décrit au paragraphe 1.2. On peut en effet l'écrire sous la forme suivante :

$$G(f_i) = 1 - \frac{1}{1 + \frac{1+\beta}{G} + j 2\pi \frac{\tau}{G} f_i}$$

On peut donc utiliser le filtre récursif mentionné précédemment en ajustant les paramètres temporels selon l'équation précédente. Pour des raisons pratiques, nous avons toutefois utilisé un filtre passe haut temporel classique, du type  $G = 1 - \frac{1}{1 + j 2\pi \tau f_i}$ , paramétré uniquement par la constante de temps  $\tau$ .

## 2.2.4 Rôle du comportement spatio-temporel de la rétine

On attribue usuellement deux intérêts fonctionnels au comportement linéaire de la rétine décrit ici.

Le premier rôle fonctionnel de la rétine serait un rehaussement de l'énergie du signal dans les fréquences hautes de l'image. On montre en effet que le spectre des images naturelles (des photographies de scènes visuelles usuelles) décroît de manière inversement proportionnelle à la fréquence considérée (Attick et Redlich, 1992, Ruderman et Bialek, 1994). Le filtre passe-bande associé à la couche plexiforme externe aurait alors pour rôle d'égaliser l'énergie dans les différentes bandes de fréquence. On appelle cela un blanchiment spectral. C'est une technique courante en traitement de l'image. Sur une image statique, cet effet correspond à faire ressortir les détails de l'image (voir fig. 2.20).

Le second rôle attribué à la rétine est la genèse des voies Parvo et Magnocellulaires, dont les spécificités semblent les destiner à des rôles fonctionnels différents. La voie Parvo serait utilisée pour la vision des détails et la reconnaissance des objets, la voie Magno pour la vision du contexte et l'analyse du mouvement. Cette séparation en différentes voies pourrait en particulier permettre de véhiculer la "bonne" information vers le "bon" centre de traitement.

Comme le lecteur le remarquera néanmoins, cette analyse est fortement teintée d'une conception passive de la vision. Nous reviendrons en particulier sur ce point au chapitre 5, où nous montrerons que l'analyse des stratégies du sujet permet de formuler d'autres hypothèses concernant le rôle fonctionnel du comportement spatio-temporel de la rétine, dans un cadre actif cette fois.

## 2.3 Dynamique des cellules de la rétine et adaptations locales

Le système visuel humain est capable de travailler dans une gamme de luminosités variant d'environ  $10^{-6}$  lux (seuil de perception) à environ  $10^4$  lux (neige au soleil) soit 10 ordres de grandeurs. On considère par ailleurs que la dynamique d'un neurone est de l'ordre de 2 ordres de grandeurs (la fréquence de décharge d'un neurone est en général comprise entre 1 et 100 Hz). Pour expliquer ce phénomène, on considère que le système visuel est capable d'adapter sa dynamique de fonctionnement en fonction de l'éclairement.

La surface de la pupille, qui constitue le diaphragme de l'œil et règle ainsi le taux de lumière incidente sur la rétine, peut varier d'un facteur 16, soit un peu plus d'un ordre de grandeur. En considérant le cas idéal où la dynamique des cônes et celle des bâtonnets seraient juxtaposées (ce

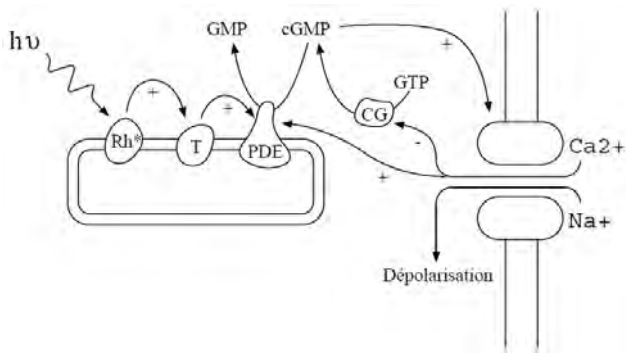


Figure 2.23 : cascade enzymatique responsable de l'ouverture des canaux ioniques lors de la stimulation du photorécepteur par la lumière. D'après Beaudot,

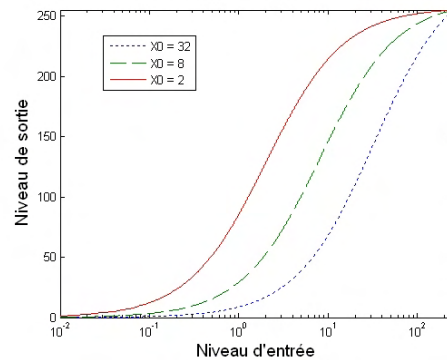


Figure 2.24: loi de Michaelis Menten pour différentes valeurs de  $X_0$

qui n'est pas vrai en pratique), et qu'elles sont toutes deux de l'ordre de 2 ordres de grandeurs, on obtient une dynamique totale d'environ 7 ordres de grandeurs. Il en reste donc encore 3. Ces derniers échelons sont gravés grâce à une adaptation de la dynamique de fonctionnement du système visuel en fonction de la luminosité ambiante. Des mesures psychophysiques corroborent cette hypothèse (Valeton et van Norren, 1983).

Dans notre modèle, cette adaptation a lieu au niveau des photorécepteurs et des cellules bipolaires. Cette adaptation est, par ailleurs, locale, c'est à dire qu'elle varie d'une cellule à une autre, en fonction de son environnement proche.

### 2.3.1 Dynamique du photorécepteur et égalisation locale de la luminance

Le réponse d'un cône à une stimulation par la lumière n'est pas linéaire. Le lecteur intéressé par une analyse fine de la réponse du cône pourra se reporter à van Hateren (2005).

Nous modélisons pour notre part cette réponse non linéaire par une loi de Michaelis-Menten. Cette loi a notamment été proposée par Shapley (1981). Par ailleurs, Beaudot (1994) montre que cette loi est compatible avec une analyse biochimique du fonctionnement du cône (illustrée figure 2.23), ainsi qu'avec les mesures psychophysiques, plus précisément avec la loi de Fechner.

Si  $Y$  est la réponse du cône à un niveau de stimulation  $X$ , cette loi s'écrit :

$$Y = \frac{X}{X + X_0}$$

où  $X_0$  est un paramètre qui modifie l'allure de la compression.

Le graphe 2.24 présente plusieurs courbes de ce type en fonction du logarithme de  $X$ , pour différentes valeurs du paramètre  $X_0$ . Cette loi est globalement linéaire pour les faibles valeurs de stimulation (i.e.  $X \ll X_0$ ), elle sature pour les fortes valeurs de stimulation (i.e.  $X \gg X_0$ ). Entre les deux, c'est à dire au voisinage de  $X_0$ , la loi de Michaelis-Menten est proportionnelle à une loi logarithmique.

Dans la plupart des modèles, le paramètre  $X_0$  est adapté en fonction de la luminance globale de l'image, ce qui permet de rendre compte des mesures psychophysiques où l'on fait varier la luminance de tout l'arrière plan. Dans notre modèle en revanche, nous adaptons le paramètre  $X_0$  à



Figure 2.25 : effet de l'adaptation locale des cônes. A gauche, image d'origine. A droite, image après une adaptation locale.

partir de la sortie des cellules horizontales. Une rétroaction de ce type est par exemple mentionnée par van Hateren (2005). Les cellules horizontales, dont la sortie correspond en quelque sorte à une moyenne spatio-temporelle locale de la sortie des photorécepteurs, permettent ainsi de réaliser une adaptation du gain de la fonction de transduction du photorécepteur en fonction de la luminance locale. A défaut d'un modèle explicite de cette rétroaction, nous utilisons une relation linéaire entre la sortie des cellules horizontales et le paramètre  $X_0$ .

#### **Implémentation :**

En pratique, les images que nous utilisons sont définies sur 255 niveaux de gris. Nous choisissons comme borne maximale pour  $X_0$  la valeur 255. La borne minimale de  $X_0$  est ajustable par l'utilisateur à l'aide du paramètre  $V_0$  qui fixe respectivement cette dernière à  $(255 - V_0)$ .

La figure 2.25 illustre l'effet de la compression adaptative locale. L'égalisation de la luminance permet de faire ressortir les détails des zones sombres. Le contraste dans les zones claires est en revanche atténué.

### **2.3.2 Dynamique des cellules ganglionnaires et égalisation locale du contraste**

La capacité d'adaptation n'est pas l'apanage des photorécepteurs. Les cellules ganglionnaires possèdent également une capacité d'adaptation (e.g. Shapley et Victor, 1979). Le point remarquable au niveau des cellules ganglionnaires est que cette adaptation a lieu séparément sur les cellules de type ON et les cellules de type OFF. Cette action séparée sur chaque canal conduit non plus à une adaptation à la luminance, mais à une adaptation au contraste.

Pour modéliser l'adaptation de la dynamique des cellules ganglionnaires, nous utilisons un modèle similaire au cas des photorécepteurs. La rétroaction est cette fois pilotée par les cellules amacrines, modélisées par un filtre spatio-temporel passe-bas du même type que les cellules horizontales. L'adaptation est effectuée séparément sur les canaux ON et OFF.

La figure 2.26 illustre l'effet de la compression adaptative au niveau des cellules ganglionnaires. Par rapport à une simple adaptation locale de la luminance, l'adaptation des cellules ganglionnaires permet de régulariser le contraste, y compris dans les zones claires.



Figure 2.26 : effet de l'adaptation locale des cônes et des cellules ganglionnaires. A gauche, image d'origine. A droite, image après adaptation locale.

### 2.3.3 Rôle de la dynamique des cellules rétiniennes

Comme nous l'avons mentionné, le rôle essentiel joué par la dynamique des cellules de la rétine est de ramener les variations du signal lumineux sur plus de 8 ordres de grandeur à des variations compatibles avec leur transmission ultérieure sur des fibres neurales. Ce rôle fonctionnel est médiatisé par deux phénomènes : la réponse non linéaire de la cellule à la stimulation et l'adaptation de cette réponse au contexte local.

Un autre aspect fonctionnel lié à la réponse non linéaire des cellules photoréceptrices mérite d'être ici mentionnée. La lumière réfléchie par un objet est en fait le produit de deux termes, l'illumination et la réflectance de l'objet. Or nous avons vu que la réponse du photorécepteur est de type logarithmique autour de son point de fonctionnement (figure 2.7). Vu les propriétés du logarithme, le signal à la sortie du photorécepteur n'est donc plus le produit de deux termes mais leur somme, ce qui rend plus aisée leur séparation.

## 2.4 Echantillonnage spatialement variant et projections corticales

On modélise usuellement la projection rétino-corticale par une projection de type Log-Polaire (Schwartz, 1977, 1980) représentée figure 2.27. Cette projection log-polaire peut être vue comme la combinaison d'une magnification corticale qui tire ses origines de la variation de la densité des cellules afférentes en fonction de l'excentricité rétinienne, et d'une projection polaire de ces cellules afférentes sur le cortex visuel primaire.

Pour des raisons de clarté, nous allons considérer séparément la question de la magnification corticale (qui correspond à la partie "log" de la projection), et la question de la projection des radiations optiques sur le cortex visuel (qui correspond à la composante "polaire" de la projection).

### 2.4.1 Modélisation de la magnification corticale

Dans notre modèle, nous considérons que la magnification corticale est uniquement liée à l'échantillonnage spatialement variant au niveau des photorécepteurs. Nous modifions donc directement l'image d'entrée selon une projection conforme aux observations de Schwartz.

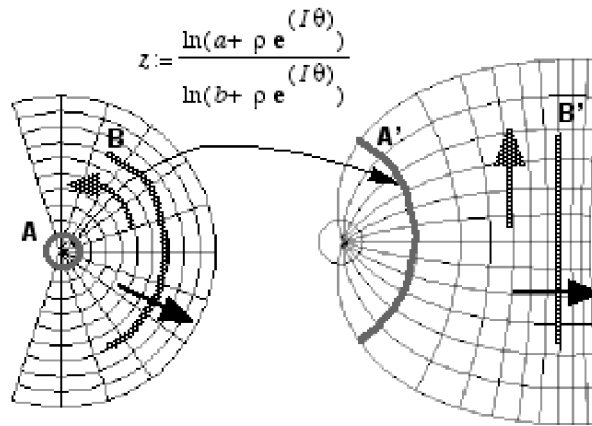


Figure 2.27 : la projection Log-Polaire. En périphérie, une expansion du champ (flèche noir) et une rotation (flèche grisée) deviennent des translations. D'après Héault, sous presse.

Nous utilisons, pour cette projection, une loi de Michaelis-Menten. Si l'on nomme  $\rho$  l'excentricité d'un cône et  $\rho'$  l'excentricité de sa projection sur le cortex (sans tenir compte des aspects polaires pour le moment), cette loi s'écrit :

$$\rho' = \rho_1 \frac{\rho}{\rho + \rho_0}$$

avec  $\rho_1$  et  $\rho_0$  deux paramètres de la projection. La figure 2.28 illustre cette projection.

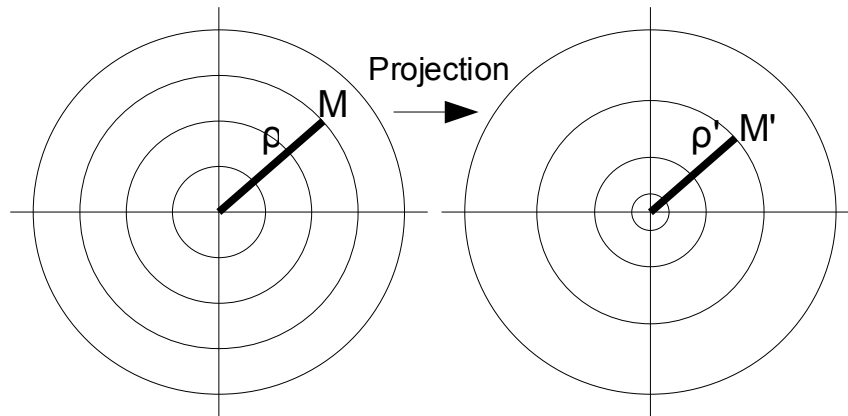


Figure 2.28 : projection logarithmique. Le point M d'excentricité  $\rho$  se projette en M', d'excentricité  $\rho'$ .

Comme nous l'avons vu, cette loi est proportionnelle à un logarithme dans une large gamme de valeurs. Par ailleurs, contrairement au logarithme, cette loi présente l'avantage d'être définie (et nulle) en 0, c'est à dire qu'elle est valable également aux faibles excentricités.

Remarque : la variation de la densité des cellules afférentes sur le cortex en fonction de l'excentricité n'est pas le seul fait des photorécepteurs. Le taux final de magnification cortical est également lié à l'arborisation dendritique des cellules bipolaires et des cellules ganglionnaires à une excentricité donnée. Dans la voie Parvo néanmoins, où un cône est contacté par une seule cellule bipolaires Midget, elle même n'étant contactée que par une seule cellule ganglionnaire Midget, ce modèle est tout à fait valable.



Figure 2.29 : application de la projection retino-corticale. A gauche : image d'origine. Au centre : filtrage passe-bas spatialement variant de l'image d'origine. A droite : image projetée, avec un taux de magnification centrale  $R0 = 4$  et un taux de réduction de l'image  $k = 1$  (pas de réduction).

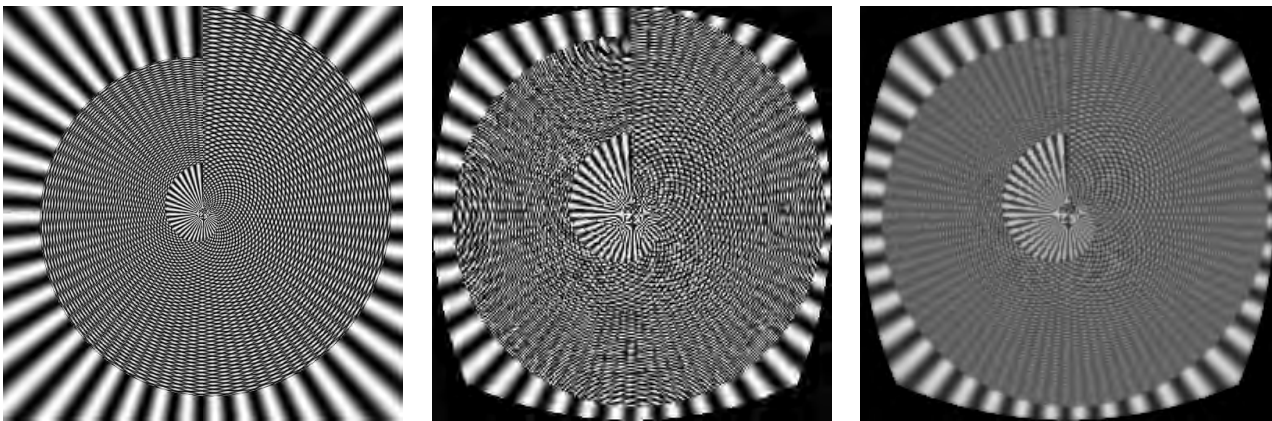


Figure 2.30: apport du filtrage passe-bas spatialement variant avant la projection. A gauche : image d'origine ; au centre : image projetée sans filtrage préalable ; à droite : image projetée après filtrage.

### Implémentation :

Pour implémenter cette projection, on procède à rebours. On considère l'image de sortie. Pour chaque pixel de cette image on calcule son excentricité  $\rho'$ . On calcule ensuite l'excentricité  $\rho$  du point correspondant dans l'image d'entrée. On reporte ensuite le niveau de gris du pixel correspondant de l'image d'entrée dans l'image de sortie.

Les paramètres  $\rho_1$  et  $\rho_0$  sont déterminés par deux conditions aux limites :

1 – Le taux  $R0$  de magnification centrale, c'est à dire le rapport entre la densité d'échantillonnage dans l'image projetée sur la densité dans l'image d'origine

2 – Le taux de réduction de l'image :  $k$ . L'image projetée peut en effet être de taille inférieure à l'image d'origine. Le rapport  $k$  est alors le rapport entre la taille d'un côté de l'image d'origine sur la taille d'un côté de l'image projetée. On considère alors que le pixel au bord de l'image projetée, réduite d'un facteur  $k$ , doit correspondre au pixel au bord de l'image d'origine. Une description détaillée de la méthode d'implémentation est fournie en annexe.

Le sous-échantillonnage de l'image pose par ailleurs un autre problème : celui du repliement spectral. Sans entrer dans les détails à ce sujet, qui seront développés au chapitre 3, mentionnons

simplement que pour effectuer le sous-échantillonnage d'une image à une fréquence donnée  $f_c$ , il faut s'assurer que celle-ci ne présente pas de fréquences supérieures à  $f_c/2$ . On utilise pour cela un filtre passe-bas qui coupe les fréquences au delà de cette fréquence limite. Dans notre cas toutefois, la fréquence d'échantillonnage est variable, il faut donc utiliser un filtre dont la fréquence de coupure est également spatialement variante. Voici comment nous avons procédé :

1. A partir de la loi de projection, nous avons calculé le taux de sous échantillonnage équivalent en fonction de l'excentricité.
2. Nous avons ensuite utilisé un filtre passe bas récursif du premier ordre (du même type que celui utilisé pour modéliser les cellules rétinienne), dont la fréquence de coupure est adaptée au taux de sous-échantillonnage calculé précédemment.

Le détail de cette étape est également fourni en annexe.

La figure 2.29 illustre l'application de la projection définie ici. La figure 2.30 montre le résultat d'un échantillonnage spatialement variant avec et sans filtrage préalable, et les artefacts liés au repliement spectral dans le premier cas.

## 2.4.2 Projection polaire sur le cortex

Le principe de la transformée polaire est très simple : il suffit de considérer la norme et l'argument d'un point dans sa représentation polaire, et d'utiliser cette même norme et ce même argument dans une représentation cartésienne. La figure 2.30 illustre cette transformation.

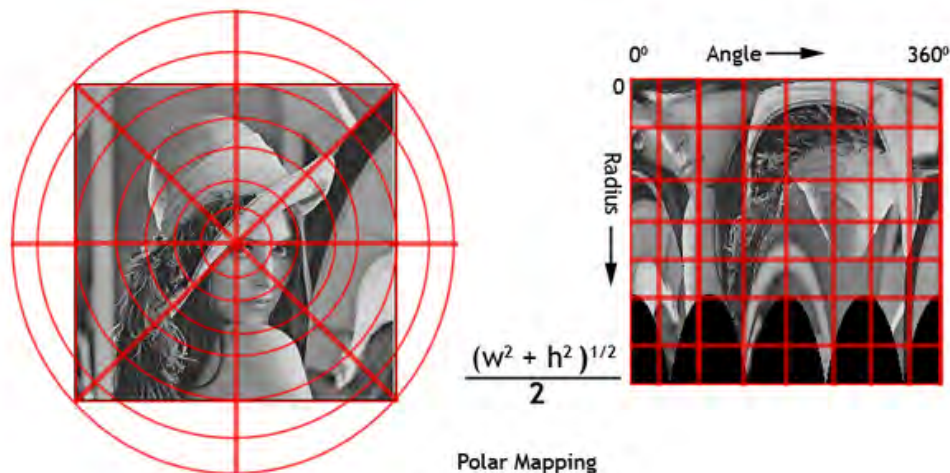


Figure 2.30 : illustration de la transformée polaire d'une image. D'après TheCodeProject, 2009.

Dans le cadre de notre étude toutefois, nous n'avons pas eu besoin d'implémenter explicitement cette transformation polaire au niveau d'une image. Cette transformation est implémentée directement dans le logiciel de substitution, en modifiant le codage sonore en fonction des coordonnées polaires dans l'image et non plus en fonction des coordonnées cartésiennes. Nous reviendrons sur ce sujet au chapitre 3.

## 2.4.3 Rôle de la projection rétino-corticale

On trouvera dans Schwartz, 1980 et 1994, une revue sur l'intérêt de la projection log-polaire du



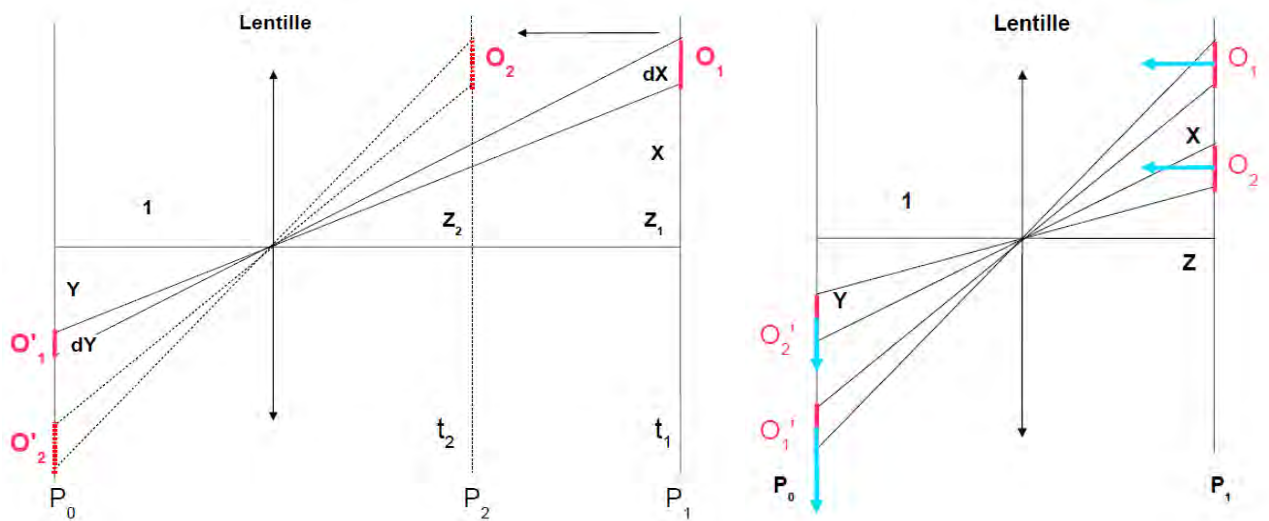


Figure 2.31: étude de l'effet d'un rapprochement de l'observateur sur la taille et la vitesse d'un objet. Sans projection logarithmique, la taille (à gauche) et la vitesse (à droite) de l'objet varie respectivement en fonction de la distance  $Z$  ou de la hauteur  $X$ .

point de vue de la perception et de la robotique. Je vais présenter ici certaines caractéristiques particulièrement pertinentes au regard d'une théorie active de la perception.

La composante polaire de la projection tout d'abord. Cette dernière transforme une opération de rotation et une opération de zoom (expansion de l'image) en deux translations orthogonales (voire figure 2.27). Cette opération change donc les relations sensorimotrices entre différents mouvements. En particulier, il peut s'avérer pertinent de pouvoir séparer facilement un mouvement d'expansion d'un mouvement de rotation. Nous reviendrons dans le chapitre 3 sur cette question et son lien avec le codage du signal au niveau de l'interface.

La composante log, ensuite, s'avère particulièrement intéressante en ce qui concerne la régularité de certaines contingences sensorimotrices visuelles.

Une prothèse visuelle, tout comme le système visuel humain, contient un dispositif optique qui projette une image sur un plan. Cette situation générique est illustrée figure 2.31. Dans cette situation, on considère un objet  $O_1$  de taille  $dX$  situé dans le plan  $P_1$  à une distance  $Z_1$  et à une hauteur  $X$  du centre du dispositif optique. Cet objet se projette sur le plan  $P_0$  (la rétine ou le capteur de la caméra) en  $O'_1$ .

Voyons maintenant les propriétés de cette projection lorsque le système optique se rapproche de l'objet. Dans ce cas, on voit que la taille de la projection de l'objet augmente (figure 2.31, à gauche).

Par ailleurs, si on considère deux objets dans le même plan frontal au dispositif et que l'on considère un mouvement d'approche de l'observateur (ou réciproquement du plan contenant les objets), on s'aperçoit que la vitesse de l'objet en approche dépend de sa hauteur dans le plan : plus l'objet est haut plus sa projection se déplace vite (figure 2.31).

Si maintenant on applique une projection de type logarithmique au plan  $P_0$ , ces deux invariances disparaissent : la taille d'un objet sur le plan de projection ne dépend pas de la distance d'observation, et la vitesse de la projection d'un objet sur ce même plan ne dépend plus de l'ordonnée du point considéré dans le plan de l'objet. De ce point de vue, une projection logarithmique régularise les contingences sensorimotrices liées à un mouvement de zoom.

**Démonstration :**

Le point de hauteur  $X$  sur le plan  $P_1$  à distance  $Z$  de la lentille se projette sur le plan  $P_0$  à la distance  $Y$  de l'axe optique telle que :

$$Y = \frac{X}{Z}$$

Si l'on applique une projection logarithmique, la nouvelle ordonnée du point sera

$$Y' = \log(Y) = \log\left(\frac{X}{Z}\right)$$

On différencie maintenant  $Y'$  :

$$dY' = \frac{dX}{X} - \frac{dZ}{Z}$$

Pour un objet contenu dans le plan  $P_1$ ,  $dZ = 0$ . Si par ailleurs  $dX$  et  $X$  (la taille et la position de l'objet) sont constantes, alors  $dY'$  (la taille de la projection de l'objet) est également constante.

Par ailleurs, si l'on dérive par rapport au temps, on obtient :

$$\frac{dY'}{dt} = \frac{1}{X} \frac{dX}{dt} - \frac{1}{Z} \frac{dZ}{dt}$$

Si l'objet est fixe dans le plan  $P_1$  (il ne se translate pas dans ce plan),  $dX/dt$  est nul, et la vitesse de  $Y'$  ne dépend plus alors de  $X$ . On pourra remarquer par ailleurs que la vitesse de déplacement du point projeté est inversement proportionnelle au "temps avant contact" avec l'objet.

## 2.5 Réalisation d'un logiciel de simulation de la chaîne rétino-corticale du primate

À l'issue de ce travail sur les différents éléments de cette chaîne rétino-corticale, sur leur modélisation et sur leur implémentation, j'ai réalisé un logiciel permettant de simuler l'ensemble de la chaîne sur des vidéos ou sur l'image en provenance d'une caméra vidéo. L'objectif de ce travail était de tester la faisabilité d'une simulation logicielle de la chaîne de traitement en temps réel pour la substitution sensorielle. Par ailleurs, nous avons fait en sorte que cet outil soit entièrement paramétrable, pour pouvoir être adapté au mieux au cas de la suppléance visuelle et pour pouvoir être utilisé à des fins de recherche ou d'application dans d'autres projets du laboratoire. Enfin, nous avons fait en sorte de regrouper l'ensemble des traitements dans un seul et même noyau, indépendant des questions d'interface utilisateur.

Le coeur de cette implémentation est un fichier C compilable sous Matlab (fichier .mex). Ce fichier contient les blocs de base du traitement rétinien en l'occurrence :

- Le filtre spatio-temporel passe bas
- L'adaptation locale
- Un filtre passe haut
- La transformée pseudo logarithmique (type Michaelis-Menten)

Ce fichier contient également une fonction de multiplexage et de démultiplexage de la couleur issue des travaux de Alleysson *et al.* (2005), qu'il eut été hors de propos de développer ici.

Ces blocs de base peuvent être agencés au gré de l'utilisateur. La figure 2.32 montre l'architecture présentée dans ce chapitre.

Une interface utilisateur générique a également été réalisée (sur la base de l'architecture ci-dessus), dans lequel l'utilisateur peut paramétrer l'ensemble des éléments de la simulation et observer leur effet sur une entrée vidéo standard (figure 2.33). En noir et blanc, le logiciel est capable de traiter environ 20 images par secondes. Ce logiciel a fait l'objet de plusieurs démonstrations notamment à NIPS (Durette *et al.*, 2006), et à l'European Retina Meeting (Durette *et al.*, 2007).

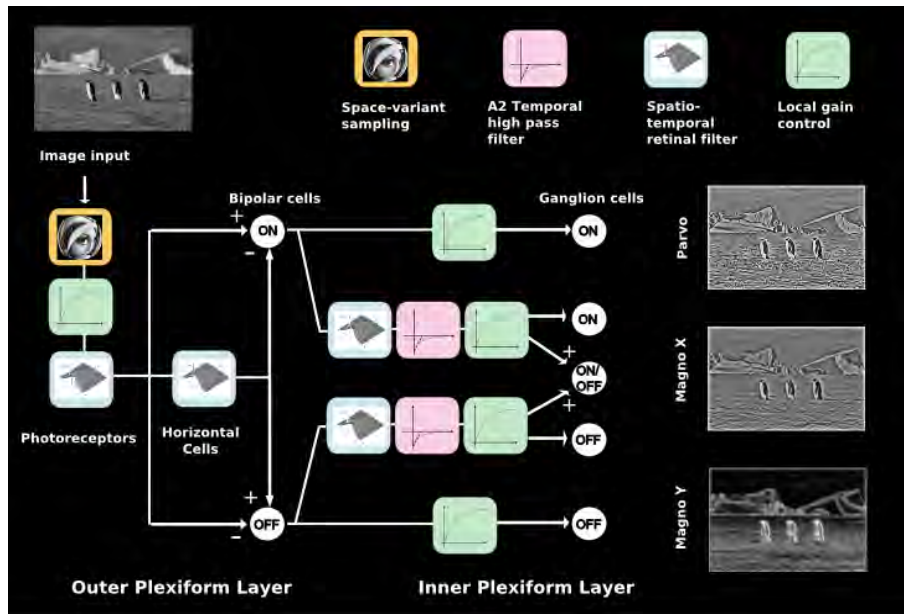


Figure 2.32 : blocs de base du logiciel et structure de la simulation.

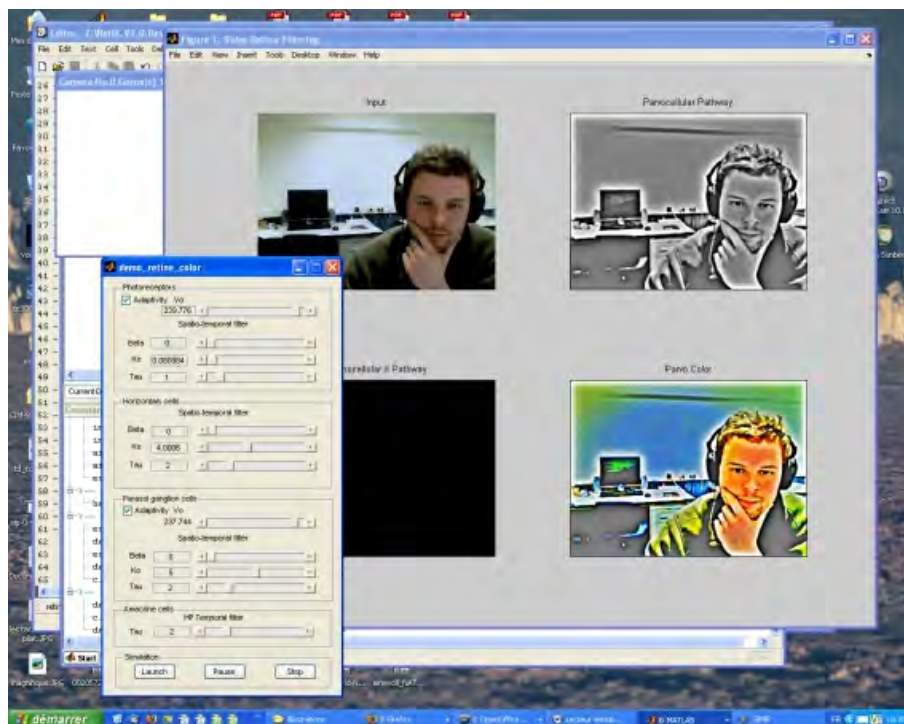


Figure 2.33 : logiciel de simulation de la chaîne de traitement rétino-cortical

## Références

- Alleysson, D.; Susstrunk, S. & Hérault, J. (2005), 'Linear Demosaicing Inspired by the Human Visual System', *Image Processing, IEEE Transactions on* **14**(4), 439--449.
- Atick, J. & Redlich, A. (1992), 'What does the retina know about natural scenes', *Neural Computation* **4**, 196B--210.
- Beaudot, W. (1994), 'Le traitement neuronal de l'information dans la rétine des vertébrés : un creuset d'idées pour la vision artificielle', PhD thesis, Institut Polytechnique de Grenoble.
- Beaudot, W.; Palagi, P. & Hérault, J. (1993), 'Realistic Simulation Tool for Early Visual Processing including Space, Time and Colour Data' International Workshop on Artificial Neural Networks', Barcelona.
- Buser, P. & Imbert, M. Hermann, ed. (1987), *Vision*.
- Carandini, M.; Demb, J. B.; Mante, V.; Tolhurst, D. J.; Dan, Y.; Olshausen, B. A.; Gallant, J. L. & Rust, N. C. (2005), 'Do we know what the early visual system does?', *J Neurosci* **25**(46), 10577--10597.
- Durette, B.; Benoit, A.; de Lavarene, B. C.; Alleysson, D. & Hérault, J. (2006), 'Real-time bio-inspired retina simulation software' Conference on Neural Information Processing Systems'.
- Durette, B.; Benoit, A.; de Lavarène, B. C.; Alleysson, D. & Hérault, J. (2007), 'A biomimetic real time retinal simulation software with non separable spatiotemporal processings, chromaticity and non linear local adaptivity', ERM'2007, Frankfurt, Germany.
- Field, G. & Chichilnisky, E. (2007), 'Information Processing in the Primate Retina: Circuitry and Coding', *Annu. Rev. Neurosci* **30**, 1B--30.
- Kolb, H.; Fernandez, E. & Nelson, R. (2009), 'Webvision. The Organization of the retina and visual system', <http://webvision.med.utah.edu/>.
- Mead, C. & Mahowald, M. (1988), 'A silicon model of early visual processing', *Neural Networks* **1**(1), 91-97.
- Mishkin, M. & Ungerleider, L. G. (1982), 'Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys.', *Behav Brain Res* **6**(1), 57--77.
- Rodieck, R. W. de boeck, ed. (2003), *La vision*.
- Ruderman, D. L. & Bialek, W. (1994), 'Statistics of natural images: Scaling in the woods', *Phys. Rev. Lett.* **73**(6), 814--817.

Schwartz, E. (1994), *Computational studies of the spatial architecture of primate visual cortex*, Springer, pp. 359--411.

Schwartz, E. (1980), 'Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding', *Vision Research* **20**(8), 645--669.

Schwartz, E. L. (1977), 'Spatial mapping in the primate sensory projection: analytic structure and relevance to perception.', *Biol Cybern* **25**(4), 181--194.

Shapley, R. M. & Victor, J. D. (1979), 'Nonlinear spatial summation and the contrast gain control of cat retinal ganglion cells.', *J Physiol* **290**(2), 141--161.

TheCodeProject (2009), "", <http://www.codeproject.com/KB/graphics/PolarView/PolarMap2.png>.

Van Hateren, H. (2005), 'A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells.', *J Vis* **5**(4), 331--347.

Wohrer, A. (2008), 'Model and large-scale simulator of a biological retina, with contrast gain control', PhD thesis, INRIA.

## Chapitre 3 : Configuration du dispositif TheVIBE

### 3.1 Les dispositifs de substitution visuo-auditive

La substitution visuo-auditive a vu le jour peu après la substitution visuo-tactile, dans les années 70. La question de la transcription de l'information visuelle en information auditive est plus délicate que dans le cas visuo-tactile. En effet, une interface tactile possède déjà une structure spatiale bidimensionnelle sur laquelle on peut tout simplement recopier l'image, ce qui n'est pas le cas a priori d'une interface sonore.

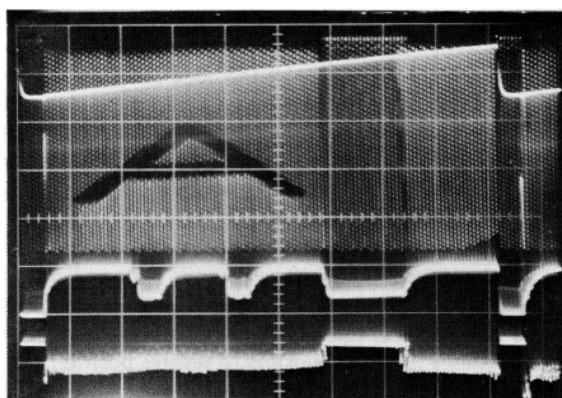


Figure 3.1 : le dispositif de Fish (1976) L'image est balayée de haut en bas et de droite à gauche par un "flying spot" (en haut). L'intensité du son est modulée par la luminosité de l'image à l'endroit où se trouve le spot à l'instant  $t$ .

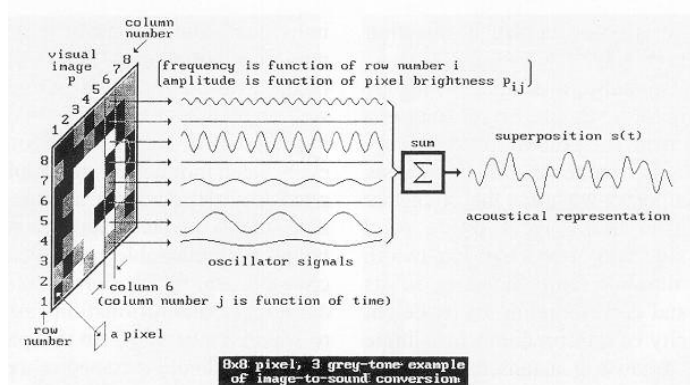


Figure 3.2 : principe de la conversion dans le dispositif The vOICE (Meijer, 1992). L'image est codée colonne par colonne. Chaque pixel d'une colonne pilote l'amplitude d'un oscillateur dont la fréquence dépend de la hauteur du pixel. Le signal généré est la somme de tous les oscillateurs.

Pour palier à cela, Fish (1976) propose de balayer l'image à l'aide d'un spot, de gauche à droite et de haut en bas à la manière d'un spot d'un tube cathodique de télévision : la composante spatiale est alors remplacée par une multiplexage temporel (fig. 3.2). L'intensité du son dépend de la luminosité dans l'image à l'endroit où se trouve le spot à l'instant  $t$ . L'image est codée en une séquence sonore.

Conscient toutefois de la difficulté qu'il pourrait y avoir à manipuler un dispositif basé uniquement sur ce multiplexage, Fish introduit d'autres indices sonores basés sur la fréquence du son et sur son panoramique (c'est à dire sur la position apparente de la source d'où provient le son). Le son, modulé en amplitude par l'intensité lumineuse à l'emplacement du spot dans l'image, est également modulé en fréquence, en fonction de la hauteur du spot dans l'image : plus le spot est haut, plus le son est aigu. Fish utilise de plus un effet de panoramique qui simule une position latérale du son perçu : quand le spot est à gauche, le son est perçu dans l'oreille gauche, puis se déplace en même temps que le spot jusque dans l'oreille droite.

L'inconvénient du dispositif de Fish est que le balayage de l'image nécessite un laps de temps qui s'accommode mal avec une utilisation en mode « camera video ». Les tests étaient d'ailleurs réalisés à partir de cassettes audio sur lesquelles la stimulation était enregistré. Malgré l'absence d'un quelconque couplage sensori-moteur dans ce cas, il réussit à obtenir que des sujets reconnaissent certains pattern sonores.

La tendance par la suite à été de présenter l'information de la manière la plus simultanée

possible, c'est à dire en réduisant le balayage. En 1992, Philip Meijer propose un nouveau dispositif de substitution visuo-auditif baptisé The vOICe (les majuscules au centre se prononçant "Oh, I see!") (Meijer, 1992). Ce dernier conserve un balayage gauche/droite de l'image, mais celle-ci est encodée colonne par colonne. Aux 64 pixels d'une colonne de l'image, le dispositif associe 64 sons purs de fréquences différentes qui sont simplement sommés, en pondérant leur amplitude par le niveau de gris du pixel correspondant. Le rythme du balayage gauche/droite est fixé à 1 Hz, le dispositif code ainsi les 64 colonnes de l'image en une seconde. Un "clic" sonore signal à l'utilisateur que l'on débute le codage d'une nouvelle image. Un effet de panoramique déplace la position apparente du signal sonore de gauche à droite à mesure que les colonnes sont encodées. Le principe de codage dans le dispositif The vOICe est illustré figure 3.2. Le dispositif de Cronly-Dillon (Cronly-Dillon, 2000) utilise le même principe de codage. Dans leur dispositif, la hauteur du pixel dans la colonne correspond à une note de musique.

Vint enfin la génération des dispositifs entièrement simultanés, c'est à dire codant chaque image sans balayage, sous la forme d'un seul son complexe. Le premier dispositif de ce type est le PSVA (Prothèse de Substitution de la Vision par l'Audition), développé par l'équipe de Claude Veraart à l'université catholique de Louvain (Capelle *et al.*, 1998). Cette fois, on attribue à chaque pixel de l'image un signal sonore qui lui est propre : la fréquence et le panoramique du son associé à ce signal est unique. Le signal sonore fourni à l'utilisateur est la somme de tous les signaux sonores associés à chacun des pixels, pondéré par leur niveau de gris respectif. Le son produit est bien sûr assez complexe, et dépend du nombre de pixels considéré dans l'image. Les auteurs choisissent donc de réduire le nombre de pixels de l'image en effectuant un sous-échantillonnage spatialement variant à la manière du système visuel (voir chapitre 2).

Le dispositif de substitution le plus récent, enfin, est le dispositif TheVIBE (Auvray *et al.*, 2005). Tout comme le PSVA, TheVIBE convertit chaque image en un son complexe à la volée. Il présente néanmoins un certain nombre d'innovations intéressantes, et en particulier une grande souplesse d'utilisation qui en fait un outil particulièrement adapté à la recherche sur ce type de prothèse. Nous allons maintenant présenter ce dispositif en détail.

## 3.2 Le dispositif TheVIBE

Le dispositif de substitution visuo-auditif TheVIBE est constitué d'une caméra, d'un PC équipé du logiciel TheVIBE et d'une paire d'écouteurs stéréophoniques. Les images en provenance de la caméra sont converties par le logiciel en sons délivrés à l'utilisateur par les écouteurs (fig. 3.3).

Le logiciel TheVIBE a été conçu par Sylvain Hanneton au laboratoire de Neurophysique et Neurophysiologie de l'université Paris V. De manière à simplifier les collaborations autour de l'évolution de ce logiciel, son concepteur a opté pour une licence libre. Le logiciel et le code source de TheVIBE sont disponible sur Internet : <http://sourceforge.net/projects/thevibe>.

### 3.2.1 Principe de la conversion visuo-sonore

Pour effectuer la conversion du flux vidéo en flux sonore, le logiciel TheVIBE utilise une analogie entre l'espace bidimensionnel de l'image et deux dimensions indépendantes de l'espace sonore : la tonalité (grave/aigüe) et le panoramique du son, c'est à dire la position dans le plan horizontal d'où le son semble provenir (fig. 3.4).

Les sons élémentaires utilisés par TheVIBE pour constituer le son complexe représentant l'image sont des sons purs (correspondant à des signaux sinusoïdaux).





Figure 3.3 : le dispositif TheVIBE. TheVIBE fonctionne sur un PC standard à la cadence de 25 images par secondes



Figure 3.4 : conversion image/son dans le logiciel TheVIBE. Un champ récepteur (cercle pointillé blanc) est un ensemble de pixels en général voisins. A chaque champ récepteur est attribué une source sonore dont les paramètres (fréquence et panoramique) dépendent de la position du champ récepteur dans l'image (matérialisée par un point blanc). Le volume sonore attribué à chaque source est fonction de la moyenne des niveaux de gris des pixels constituant le champ récepteur. Le son associé à l'image est la somme des sources sonores.

L'image est segmentée en champs récepteurs dont le nombre est configurable par l'utilisateur. Un champ récepteur est un groupe de pixels, en général voisins (fig 3.4, cercle pointillé). A chaque champ récepteur est attribué un son élémentaire. Les propriétés de ce son élémentaire (tonalité et panoramique) dépendent de la position du champ récepteur dans l'image selon une loi définie par l'utilisateur. Le son complexe délivré à l'utilisateur est la somme de tous les sons élémentaires pondérés, linéairement ou selon une loi plus complexe, par la moyenne des niveaux de gris (et donc de la luminosité) dans le champ récepteur associé.

A chaque image est donc associé un son complexe. La durée de présentation de ce son complexe est égale à la période de rafraîchissement de l'image vidéo, fixée usuellement à 25 images par secondes. La cadence visuelle et la cadence sonore sont donc égales, c'est pourquoi on parlera de la conversion d'un flux vidéo en un flux audio.

Une démonstration du fonctionnement du logiciel sur un film est disponible sur Internet : [www.lis.inpg.fr/pages\\_perso/durette](http://www.lis.inpg.fr/pages_perso/durette). J'ai également réalisé une documentation du logiciel disponible sur Sourceforge : <http://thevibe.wiki.sourceforge.net/Aide>

L'une des innovations majeures du dispositif est de ne pas travailler sur la totalité des pixels de l'image mais seulement sur des sous-ensembles regroupés en champs récepteurs. Cela permet de réduire considérablement le nombre de pixels à traiter. Par exemple, si l'on choisit d'utiliser 200 champs récepteurs, chacun constitué de 10 pixels, on travaille sur 2000 pixels, à comparer aux 76800 pixels d'une image de webcam standard (320x240). C'est en grande partie pour cette raison que le logiciel TheVIBE est capable de transcrire en temps réel le flux de la caméra en un flux audio. Toutefois, ce type de fonctionnement impose une réflexion sur la manière de répartir ces champs récepteurs dans l'image et sur leur taille. Nous aborderons en détails cette question dans le paragraphe "Echantillonnage de l'image".

L'autre innovation du dispositif est que ces champs récepteurs sont définis par l'utilisateur dans



un fichier externe, ainsi que les propriétés des sons élémentaires qui leur sont respectivement associés. La conversion entre l'espace de l'image et l'espace sonore est donc entièrement configurable par l'utilisateur. Nous aborderons ces questions dans la partie "Codage sonore".

Dans un premier temps, nous allons voir concrètement comment on configure le logiciel TheVIBE.

### 3.2.2 Configuration de la conversion visuo-sonore

La configuration de la conversion visuo-sonore dans TheVIBE se fait à l'aide d'un fichier externe au format texte. Ce fichier regroupe la définition des champs récepteurs et les propriétés des sources sonores qui leur sont respectivement associées.

La première ligne du fichier indique le nombre total de champs récepteurs utilisés, les lignes suivantes déterminent les propriétés de chacun de ces champs récepteurs. Pour illustrer le propos, un exemple de fichier de configuration est présenté ci-dessous (fig. 3.5). Il correspond à une configuration utilisant trois champs récepteurs.

La définition d'un champ récepteur se fait de la manière suivante:

1. Sur la première ligne, on indique dans l'ordre le nombre de pixels constituant le champ récepteur (ici 5 pixels), le nombre de paramètres attribués à la source sonore (ici 4), puis les paramètres de la source sonore, en l'occurrence sa fréquence (par exemple 1574Hz), son déphasage (2.4), son panoramique d'amplitude (ILD), défini entre 0 et 1, et enfin, sa différence de temps inter-aural (ITD). Nous reviendrons sur la signification de ces paramètres au paragraphe "Echelles sonores".
2. Sur les lignes suivantes, on indique les coordonnées des pixels constituant le champ récepteur.

L'analogie entre l'opération d'échantillonnage réalisée par le biais de ce fichier de configuration et l'échantillonnage de l'image opéré dans l'oeil par les photorécepteurs nous a conduit à utiliser le terme de "rétine" pour qualifier ce fichier de configuration visuo-sonore du dispositif TheVIBE. L'analogie néanmoins ne dépasse pas ce stade.

S'il est envisageable de construire manuellement un fichier de configuration pour trois champs récepteurs, la tâche devient inabordable pour 200 champs récepteurs (l'ordre de grandeur usuel pour les "rétines" utilisées dans la pratique). Les rétines sont donc générées automatiquement à partir d'un certain nombre de paramètres, notamment :

- Le nombre "N" de champs récepteurs dans l'image
- Le rayon "r" d'un champ récepteur permettant de déterminer la surface qu'il couvre au niveau de l'image
- Le nombre "n" d'échantillons dans chaque champ récepteur.

Nous aborderons en détail les méthodes permettant de générer les fichiers de configuration dans la partie "Echantillonnage de l'image", ainsi que les critères permettant de choisir le rayon et le nombre d'échantillons par champ récepteur. La question du nombre de champs à utiliser sera discutée dans la partie "Codage sonore".

Nombre de champs récepteurs						
3		Fréquence	Phase	IAD	ITD	
5	4	1574.813418	2.400115	0.883266	494.197892	} Champ récepteur n° 1
305	149					
265	96					
254	136					
284	97					
282	142					} Champ récepteur n° 2
5	4	2298.765589	4.351668	0.434281	-108.512463	
110	213					
166	195					
167	202					
141	166					} Champ récepteur n° 3
160	154					
5	4	2524.515106	4.671318	0.167535	-457.741806	
49	192					
61	217					
78	207					
57	222					
57	201					

Figure 3.5 : fichier de configuration dans TheVIBE. Exemple du fichier "retine\_3.txt". La première ligne du fichier indique le nombre totale de champs récepteurs. La suite du fichier est dédié à la définition de chaque champ récepteur. Pour chaque champ récepteur, la première ligne correspond aux propriétés de la source sonore associée, les lignes suivantes indiquent les coordonnées des pixels définissant le champ récepteur.

### 3.2.3 Les autres paramètres du logiciel

Un fichier d'option permet de définir les autres paramètres du logiciel, notamment :

- La taille de l'image et la fréquence de rafraichissement vidéo
- La taille du buffer des échantillons sonores pour chaque image ainsi que la fréquence d'échantillonnage du son
- Le pas d'enregistrement de l'activité des champs récepteurs. Dans certain cas en effet, il peut s'avérer intéressant d'enregistrer les valeurs calculées par le logiciel concernant "l'activité" des champs récepteurs. Cela permet de visualiser *a posteriori* la stimulation du point de vue du sujet. Cette "activité" est calculée à partir de la moyenne des niveaux de gris des pixels choisis pour chaque champ récepteur. Nous reviendrons sur cette question par la suite.
- Le niveau de seuil en dessous duquel l'activité d'un champ récepteur est considérée comme nulle et n'intervient donc pas dans le calcul du son final.

Un exemple de fichier d'option est disponible en annexe.

## 3.3 Echantillonnage de l'image

La taille de l'image provenant d'une caméra standard est de l'ordre de 320x240 soit 76800 pixels. Si ce nombre reste faible par rapport aux 7 millions de cônes dans la rétine, il est potentiellement élevé relativement à deux données :

1. La résolution du système auditif humain : dans quelle mesure l'oreille est-elle capable de discriminer la fréquence et le panoramique de 76800 sons élémentaires? En effet, la bande passante, c'est à dire la gamme des valeurs utiles, en fréquences et en panoramique de

l'oreille est finie : l'oreille entend des sons allant de 20Hz à 20kHz et le codage du panoramique va de +90° (son provenant de la gauche) à -90° (son provenant de la droite). Dans la mesure où la bande passante de l'oreille est finie, la capacité de transfert de l'oreille (ou encore le nombre maximale de sons qu'elle est capable de discriminer) dépend de sa résolution, c'est à dire de sa capacité à différencier deux fréquences ou deux valeurs de panoramiques.

2. La capacité de calcul du support informatique : dans la mesure où l'on souhaite coder l'information à la cadence vidéo, le calcul du son complexe associé à chaque image doit se faire en un temps limité, ce qui réduit la marge des opérations possibles.

En utilisant des champs récepteurs préalablement définis, le logiciel TheVIBE propose une solution astucieuse à ce problème : le nombre de pixels à traiter est configurable par l'utilisateur, ainsi que leur répartition et leur organisation. Néanmoins, dès lors que le nombre de pixels choisi est inférieur au nombre de pixels de l'image, on se trouve dans un cas de sous-échantillonnage. Le problème du sous-échantillonnage est bien connu dans le domaine du traitement du signal.

Dans cette première partie, nous allons étudier les problèmes qui peuvent se poser lorsque l'on sous-échantillonne un signal, ainsi que les moyens de les éviter. Nous en déduirons des méthodes permettant d'optimiser la répartition des champs récepteurs et de déterminer la distribution et le nombre de pixels qui les constituent.

### 3.3.1 Sous-échantillonnage : ce que dit le traitement du signal

Contrairement à ce que l'on pourrait croire, le sous-échantillonnage d'un signal ne conduit pas nécessairement à une perte d'information. Tout dépend en fait du contenu fréquentiel du signal, appelé également bande passante, et de la fréquence (du taux) d'échantillonnage.

#### **Le théorème de Nyquist Shannon**

Les travaux de Nyquist et Shannon ont eu pour but de déterminer dans quelle mesure un signal analogique pouvait être échantillonné de manière à ce que le signal d'origine puisse être intégralement reconstitué à partir de ses échantillons sans qu'il y ait perte d'information.

Dans le cas d'un signal unidimensionnel échantillonné régulièrement, le théorème de Nyquist-Shannon énonce que cela est possible si l'extension en fréquence du signal analogique est finie et si la fréquence d'échantillonnage est égale ou supérieure au double de la fréquence maximale contenue dans le signal:

$$f_e \geq 2 \times f_{max}$$

Dans le cas contraire, on observe ce que l'on appelle un repliement spectral. En pratique, les hautes fréquences du signal sont interprétées (et reproduites) comme des basses fréquences. L'illustration 3.6 permet de comprendre de manière intuitive ce qui se produit. Dans le cas des images cela peut produire des effets dits de moiré (fig. 3.7).

Un raisonnement simple permet de retrouver rapidement les résultats de Nyquist et Shannon. L'opération d'échantillonnage usuelle peut être modélisée par la multiplication du signal par ce qu'on appelle un peigne de Dirac, c'est à dire une somme de Diracs disposés à intervalles réguliers de distance  $\Delta$  (fig. 3.8). Soit  $\hat{e}$  le signal échantillonné,  $e$  le signal d'origine, et  $\Delta$  le pas d'échantillonnage, on peut écrire :

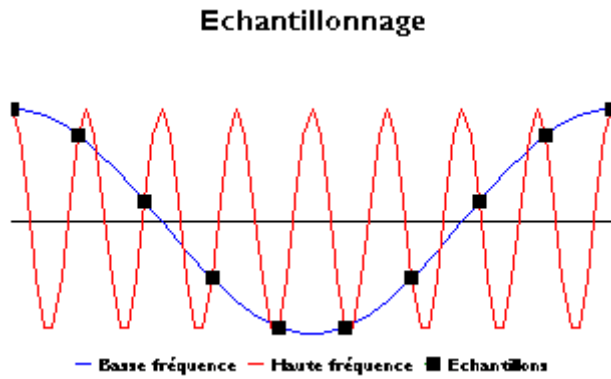


Figure 3.6 : phénomène de repliement spectral. Dans le cas où la fréquence d'échantillonnage ne vérifie pas les conditions de Nyquist-Shannon, le signal reconstitué à partir des échantillons correspond à un décalage en basse fréquence du signal d'origine.

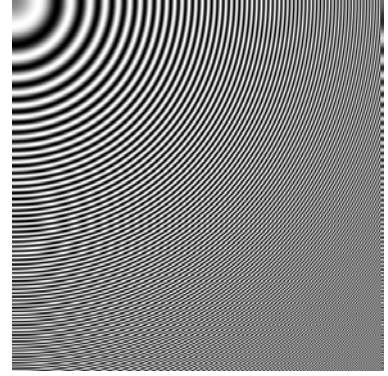


Figure 3.7 : effet du repliement spectral sur une image réduite de moitié. L'image d'origine est constituée du quart d'une image constituée de cercles concentriques. Le sous-échantillonnage produit des artefacts en basse fréquence (phénomène de moiré)

$$\hat{e}(x) = e(x) \times \sum_k \delta(x - k \cdot \Delta)$$

Prenons la transformée de Fourier de cette équation. La transformée de Fourier d'un peigne de Dirac de paramètre  $\Delta$  est un nouveau peigne de Dirac de paramètre  $1/\Delta$ . Par ailleurs, la multiplication devient, dans l'espace de Fourier, un produit de convolution. On peut donc écrire :

$$\hat{S}(v) = S(v) * \sum_k \delta(v - k \cdot \frac{1}{\Delta})$$

L'opération d'échantillonnage dans le domaine spatial correspond donc, dans le domaine de Fourier, à une réplique du spectre du signal à intervalles réguliers de pas  $1/\Delta$  (fig. 3.9). Pour reconstituer le signal, il suffit d'isoler l'une de ces copies dans le domaine spectral. Cette opération est facile à réaliser, à condition que les répliques spectrales du signal ne se juxtaposent pas. Cette condition est réalisée si et seulement si l'extension en fréquence du signal échantillonné ne dépasse pas  $1/(2\Delta)$ , c'est à dire la moitié de la fréquence d'échantillonnage.

### Extension au cas bidimensionnel :

Dans le cas usuel, la grille d'échantillonnage d'une image est régulière et à maille carrée. Les résultats obtenus dans le cas unidimensionnel sont alors facilement généralisables. En effet, l'échantillonnage selon une grille régulière à maille carrée correspond à la multiplication de l'image par une "brosse de Dirac". Une brosse de Dirac correspond à la convolution de deux peignes de Dirac orthogonaux.

Dans ce cas, l'opération d'échantillonnage sur une grille à maille carrée de côté  $\Delta$  conduit, dans le domaine de Fourier, à la réplique du spectre sur les noeuds d'une grille à maille carrée de côté  $1/\Delta$ . L'effet du sous-échantillonnage d'une image dans le domaine spatial et dans le domaine spectral est présenté fig. 3.10. De la même manière que dans le cas unidimensionnel, l'effet de repliement spectral apparaît si la fréquence d'échantillonnage selon chacun des axes est inférieure à  $1/(2\Delta)$ .

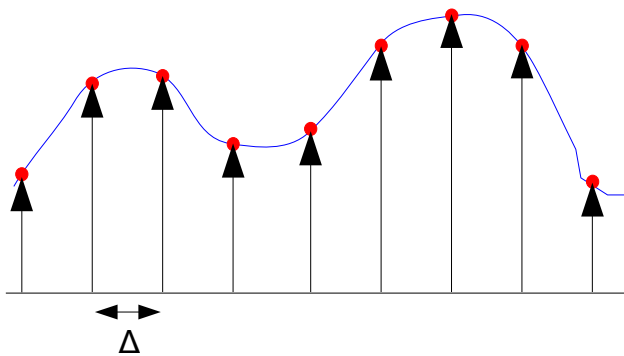


Figure 3.8: échantillonnage d'un signal L'opération d'échantillonnage correspond à la multiplication du signal (en bleu) par un peigne de Dirac (flèches noires).

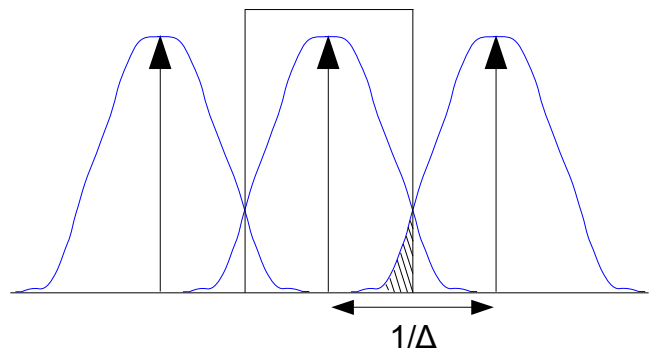


Figure 3.9 : effet de l'échantillonnage dans le domaine de Fourier. L'opération d'échantillonnage produit une réplique du spectre du signal à intervalle régulier. Si l'extension du spectre original est supérieure à la moitié de la fréquence d'échantillonnage, les copies du spectre se mêlent avec le spectre d'origine (partie hachurée)



Image d'origine

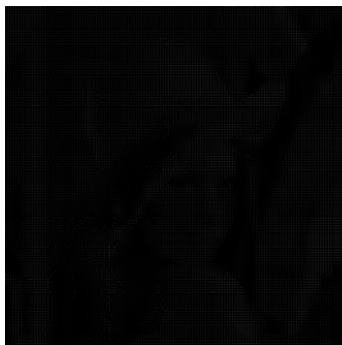
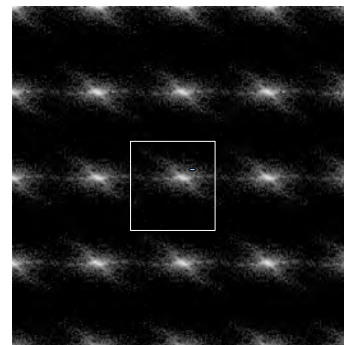


Image sous-échantillonnée (1/4)



Spectre de l'image d'origine



Spectre de l'image échantillonnée

Figure 3.10 : échantillonnage d'une image vue dans le domaine spatial et le domaine spectral Le sous-échantillonnage correspond à une réplique du spectre dans le domaine de Fourier. Tant que les spectres ne se superposent pas, le signal peut être reconstitué fidèlement à partir de ses échantillons. Le spectre doit donc être contenu dans un carré de taille  $1/(2\Delta)$  de côté. Dans le cas présent, le nombre d'échantillons a été réduit par 9. Dans le cas du dispositif TheVIBE, le taux de sous-échantillonnage est de l'ordre de 384!

### S'affranchir du repliement spectral

Il y a deux possibilités pour s'affranchir d'un problème de repliement spectral :

- 1 - Augmenter le taux d'échantillonnage.
- 2 - Supprimer du signal original les fréquences au delà de la fréquence de Nyquist-Shannon.

Dans notre cas, le nombre d'échantillons est fixé par l'utilisateur : il s'agit du nombre de champs récepteurs. La méthode que nous allons employer consiste donc à filtrer l'image d'entrée de manière à ce que son extension en fréquence soit adaptée à un sous-échantillonnage par  $N$  points. Concrètement, il s'agit de réaliser un filtrage "passe-bas" qui coupe les hautes fréquences du signal et laisse passer les basses fréquences. Ce filtrage sera réalisé par l'opération de moyenne sur les échantillons du champ récepteur. Nous reviendrons en détail sur cette question.

Par ailleurs, l'utilisation d'une grille d'échantillonnage différente de la grille à maille carrée traditionnelle ou bien l'ajout d'un composante aléatoire sur la position de l'échantillon (couramment appelé jitter) permettent de limiter les effets du repliement spectral. Nous reviendrons également sur ces questions par la suite.

Nous allons voir maintenant comment ces connaissances peuvent être appliquées concrètement pour générer un fichier de configuration pour TheVIBE.

### 3.3.2 Création de rétines par échantillonnage aléatoire uniforme

Livré avec le logiciel TheVIBE, un programme fonctionnant sous Scilab permet de générer automatiquement un fichier "rétine" dont le nombre  $N$  de champs récepteurs, le rayon  $r$  de ces champs récepteurs et le nombre  $n$  d'échantillons choisis dans chaque champ récepteur est déterminé par l'utilisateur. L'élément original de ce programme est de générer ces champs récepteurs à l'aide de fonctions aléatoires.

La génération du fichier de configuration se fait de manière progressive, champ récepteur par champ récepteur selon l'algorithme suivant, illustré figure 3.11.

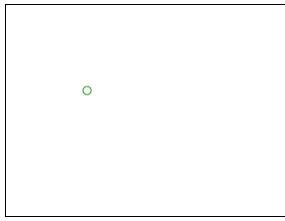
Etape 1 : on choisi aléatoirement une position dans l'image, selon une loi uniforme. Cette position constitue le centre du champ récepteur.

Etape 2 : on choisi, toujours selon une loi aléatoire de répartition uniforme,  $n$  échantillons dans une région carrée de coté  $2r$  autour de ce point.

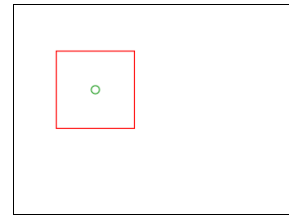
Ces deux opérations sont réitérées  $N$  fois.

Le résultat obtenu pour  $N = 200$ ,  $r = 10$ ,  $n = 10$ , conformément à une étude précédente réalisée sur le dispositif (Auvray 2005), est présenté sur la figure 3.12.

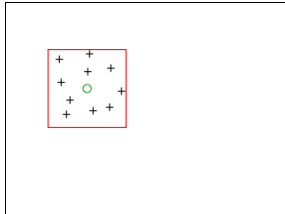
On constate que les rétines générées de cette manière sont très irrégulières. Elles présentent des zones où la densité des champs récepteurs et des échantillons est très forte, et des zones où la densité des champs récepteurs est nulle : l'image n'est pas échantillonnée à ces endroits. Par ailleurs, cette non-stationnarité de la densité des champs récepteurs limite toute tentative d'approche théorique concernant les problèmes liés au repliement spectral.



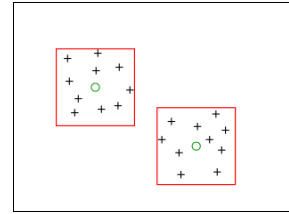
Choix du centre du champ récepteur



Zone couverte par le champ récepteur



Choix des n points dans le champ récepteur



Itération de l'opération

Figure 3.11 : génération du fichier de configuration par échantillonnage aléatoire uniforme

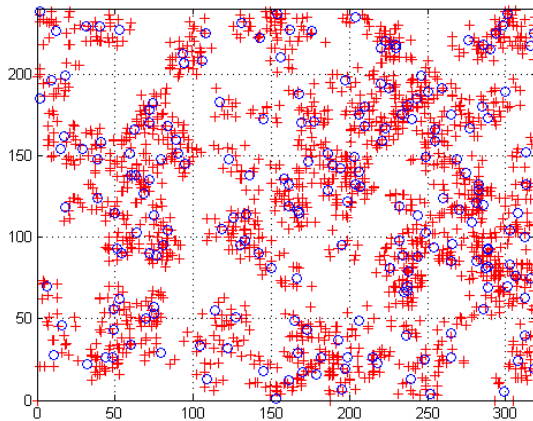


Figure 3.12 : rétine obtenue par la méthode d'échantillonnage aléatoire uniforme pour  $N=200$ ,  $r=10$  et  $n=10$ . Les cercles représentent le centre des champs récepteurs, les croix sont les échantillons. La distribution des champs récepteurs est très hétérogène, tout comme celle des échantillons.

Il convient ici de préciser que, dans le cadre sensorimoteur, ces défauts ne sont pas rédhibitoires. On pourra se reporter par exemple à la partie concernant la tâche aveugle dans l'article de O'Regan et Nöé (2001), qui correspond à une question équivalente. Ce genre de défaut reste toutefois susceptible de diminuer l'efficacité du dispositif. Dans une tâche de localisation, notamment, dont le détail sera présenté dans le chapitre 4, cela provoquait une perte fréquente du contact sensoriel avec la source lumineuse, et gênait l'apprentissage des lois de covariation sensori-motrices associées au dispositif. Nous avons donc entrepris d'améliorer ce générateur de rétine afin d'obtenir une meilleure couverture de l'image.

### 3.3.3 Création de rétines par division de l'image en rectangles élémentaires

Dans un premier temps, nous avons découpé l'image en  $N$  rectangles élémentaires ( $N$  étant le nombre de champs récepteurs). Dans chacun de ces rectangles, on choisit un point qui constitue le centre du champ récepteur : par rapport à une distribution totalement aléatoire, on améliore de cette manière la répartition des champs récepteurs. Les pixels attribués à chaque champ récepteur sont alors choisis comme auparavant dans un carré de côté  $2r$  autour de son centre.

**Choix du paramètre  $r$**  : le rayon  $r$  du champ récepteur est choisi égal à la taille du côté le plus long de chaque rectangle élémentaire. De la sorte, quelque soit la position des centres des champs récepteurs dans deux boîtes voisines, on est sûr que les champs récepteurs se superposent. De cette manière, on limite les discontinuités, les "trous", dans l'échantillonnage.

La figure 3.13 illustre le résultat obtenu avec cette méthode. L'image de 320 par 240 pixels est segmentée en  $16 \times 12$  carrés de 20 pixels de côté. La distance maximale entre les centres de deux champs récepteurs est donc de 40 pixels. Pour assurer une continuité de l'échantillonnage, les échantillons sont choisis dans un carré de 40 pixels de côté ( $r=20$ ).

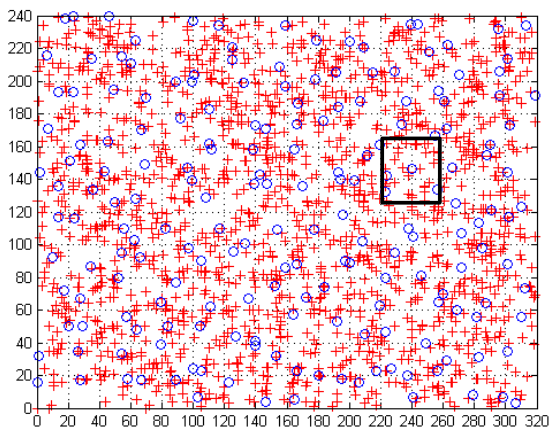


Figure 3.13 : rétine obtenue par la méthode de division de l'image en rectangles élémentaires, ici  $N=192$ ,  $r=20$  et  $n=10$ . Les rectangles élémentaires sont représentés en pointillé. Le carré noir illustre l'extension d'un champ récepteur.

On constate que la répartition des champs récepteurs est plus régulière et que la couverture de l'image semble moins parsemée. Dans la pratique, ces rétines ont été utilisées dans une expérience de localisation présentée au chapitre 4.

Néanmoins, la répartition des champs récepteurs reste encore relativement irrégulière, en particulier, lorsque l'on utilise peu de champs récepteurs : les rectangles élémentaires utilisés pour diviser l'image en régions sont alors de taille importante et la disposition des champs récepteurs moins homogène.

Par ailleurs, le critère utilisé concernant la taille des champs récepteurs correspond à une valeur maximale, déterminée de manière à être sûr du recouvrement entre deux champs récepteurs voisins. De ce fait, l'espace échantillonné par le champ récepteur est particulièrement grand, et cela nuit à la résolution spatiale du dispositif. De plus, l'échantillonnage d'un espace large nécessite de nombreux échantillons, nous reviendrons sur ce point par la suite. Nous allons donc voir maintenant une méthode permettant de positionner de manière optimale les champs récepteurs dans l'image, et de déterminer un critère théorique permettant de choisir la taille du champ récepteur et le nombre d'échantillons dans chaque champ récepteur.



### 3.3.4 Création de rétines par auto-organisation de Kohonen

L'idée ici est d'utiliser une méthode d'auto-organisation bien connue en sciences cognitives de manière à ce que les champs récepteurs se répartissent de manière uniforme sur toute l'image. La densité des champs récepteurs sera alors stationnaire, ce qui permettra d'utiliser le théorème de Nyquist-Shannon pour obtenir un critère concernant la taille des champs récepteurs.

#### 3.3.4.1 Distribution des champs récepteurs

L'objectif ici est de répartir de manière homogène les N centres des champs récepteurs dans l'image. Pour cela, nous utilisons une version simplifiée de l'algorithme de Kohonen décrite dans Hérault et Jutten, (2001), p177. Cet algorithme est à l'origine un algorithme de quantification vectorielle : pour un jeu de donnée fixé, il permet de déterminer les vecteurs permettant de représenter au mieux ces données. Dans le cas présent, nous allons appliquer cet algorithme au centre des champs récepteurs en simulant un choix aléatoire uniforme de pixels dans l'image. Les centres des champs récepteurs se répartiront alors au mieux pour représenter cette distribution, c'est à dire de manière homogène.

##### **Principe de l'algorithme :**

- 1- On choisit un ensemble C de N points répartis aléatoirement dans la zone à échantillonner.
- 2- On tire aléatoirement un point P selon une loi aléatoire uniforme dans la zone à échantillonner,
- 3- On identifie le point M de l'ensemble C le plus proche du point P,
- 4- On modifie le point M en M' selon la loi d'adaptation :

$$M' = M - \mu * (M-P)$$

- 5- On teste une condition d'arrêt. Si le test est négatif, on reprend en 2.

##### **Convergence et condition d'arrêt :**

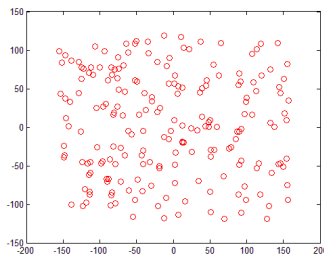
Pour faciliter la convergence, le paramètre  $\mu$  varie au fur et à mesure des itérations. Toujours selon Hérault et Jutten, une évolution du pas d'adaptation en  $1/k$  (k étant le compteur d'itération) permet d'obtenir une convergence presque sure (ps). Concrètement, nous avons choisi pour  $\mu$  une expression de la forme :

$$\mu(k) = \mu_0 / (1 + \alpha * k)$$

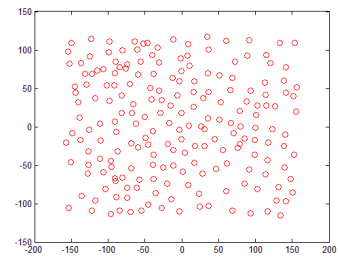
Pour vérifier la convergence de l'algorithme, on calcule la moyenne de la distance entre le point M et le point P sur 1000 itérations. Au fur et à mesure de l'homogénéisation de la distribution, la distance entre les points choisis aléatoirement et leur champ récepteur respectivement le plus proche diminue.

##### **Résultats :**

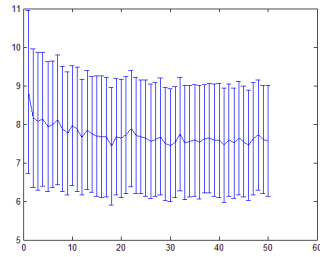
La figure 3.14 illustre le fonctionnement et les résultats de l'algorithme pour répartir le centre de 200 champs récepteurs dans un rectangle de 320 x 240 pixels. Les paramètres de convergence, déterminés empiriquement sont les suivants :  $\mu_0 = 0.8$ ,  $\alpha = 0.0005$ . En considérant l'évolution du critère de convergence décrit précédemment, la condition d'arrêt a été fixée de manière empirique à 50000 itérations.



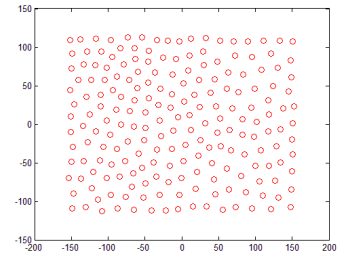
Distribution initiale



Distribution après 1000 itérations



Evolution du critère d'erreur



Distribution finale

Figure 3.14 : répartition des champs récepteurs par la méthode de Kohonen

A la sortie de l'algorithme, la distribution des champs récepteurs est uniforme et leur densité stationnaire. Soit  $A$  l'aire échantillonnée, on peut déduire la densité des champs récepteurs :

$$\rho = \frac{N}{A}$$

ainsi que la distance moyenne approximative au premier voisin, séparant donc deux champs récepteurs :

$$d = \sqrt{\left(\frac{1}{\rho}\right)}$$

A.N. :  $A = 324 \times 240 = 76800 \text{ p}^2$  et  $N=200$  champs récepteurs  $\Rightarrow d \sim 20 \text{ pixels}$

### 3.3.4.2 Répartition des échantillons dans les champs récepteurs

Maintenant que la question de l'homogénéité de la distribution des champs récepteurs est réglée, on peut aborder les questions présentées plus haut concernant le problème du repliement spectral.

#### Problématique :

On reçoit une image issue de la caméra de taille  $H \times W$  pixels que l'on souhaite sous-échantillonner sous la forme de  $N$  valeurs issues d'une opération de moyenne sur  $N$  groupes de  $n$  pixels appelés champs récepteurs et définis au préalable.

Ce sous-échantillonnage nécessite un filtrage passe-bas sur l'image d'entrée. Dans quelle mesure et sous quelles conditions l'opération de moyenne sur les échantillons de chaque champ récepteur peut-elle être assimilée à un filtrage passe-bas satisfaisant au regard de la question du repliement spectral? Il s'agit en particulier de déterminer la répartition de ces pixels dans les champs récepteurs, ainsi que leur nombre.

### Modélisation du problème :

Dans la mesure où le nombre  $H \times W$  de pixels dans l'image d'entrée (76800 pour une image standard) est très important comparé au nombre  $N$  de valeurs dans l'image de sortie (le nombre de champs récepteurs, de l'ordre de 200, à ne pas confondre avec le nombre  $n$  d'échantillons dans chaque champ récepteur), on se place dans le cas de l'échantillonnage (discret) d'un signal continu.

Nous avons vu, au paragraphe 2.1, que le sous-échantillonnage d'une image ne produisait pas d'artefact si la bande passante du signal était inférieure à une fréquence limite. Dans notre cas, nous souhaitons appliquer à l'image un filtre qui atténue les fréquences au delà de cette fréquence limite. Un filtre passe-bas est en général caractérisé par une fréquence au delà de laquelle on considère que le signal est atténué. Cette fréquence s'appelle la fréquence de coupure. Nous devons donc réaliser un filtre passe-bas dont la fréquence de coupure est égale à la fréquence limite déterminée par le sous-échantillonnage.

### Détermination de la fréquence limite associée au sous-échantillonnage :

La fréquence limite associée au sous-échantillonnage de l'image dépend d'une part de la distance moyenne " $d$ " entre les champs récepteurs mais aussi de leur organisation spatiale.

En première approximation, on peut considérer que les champs récepteurs sont organisés selon une maille régulière carré de côté " $d$ ". Dans ce cas, comme nous l'avons vu au §2.1, la fréquence limite associée est:

$$f_c = \frac{f_e}{2} = \frac{1}{2d}$$

A.N. :  $d = 20 \text{ pixels} \Rightarrow f_c = 0.026 \text{ pixels}^{-1}$

Toutefois, ce modèle représente mal l'organisation effective des champs récepteurs. Un modèle plus conforme à la réalité est celui d'un échantillonnage à maille hexagonale centrée (ou maille triangulaire). Ce type de maillage est bien connu en chimie car sa compacité est optimale.

La transformée de Fourier d'un réseau hexagonal centré de paramètre " $a$ " donne un réseau hexagonal centré de paramètre  $\frac{2}{\sqrt{3}}a$  (fig. 3.15). La fréquence limite dans notre cas vaut donc :

$$f_c = \frac{1}{\sqrt{3}d}$$

A.N. :  $d = 20 \text{ pixels} \Rightarrow f_c = 0.029 \text{ pixels}^{-1}$

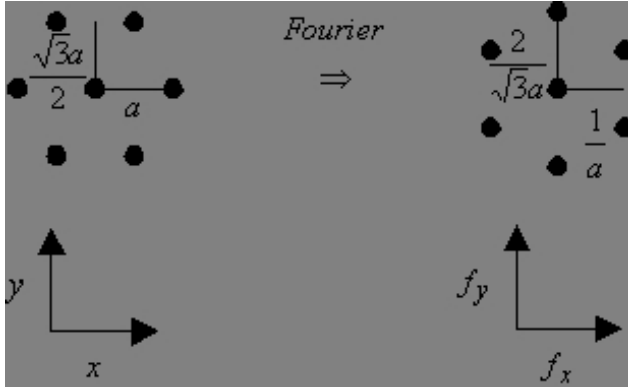


Figure 3.15 : transformée de Fourier d'un maillage hexagonal centré

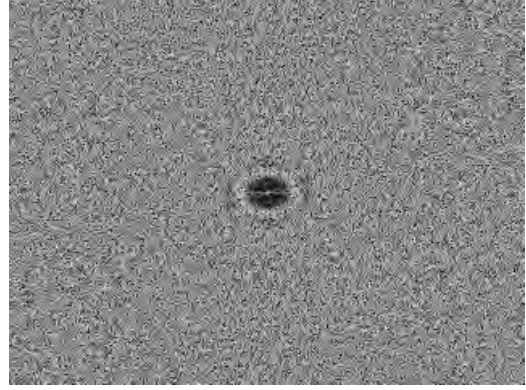


Figure 3.16 : transformée de Fourier de la répartition des champs récepteurs dans l'image (N=200)

Nous avons enfin évoqué dans le §2.1 l'intérêt des variations aléatoires de la position des échantillons autour des noeuds du réseau d'échantillonnage, appelé communément "jitter". Le lecteur intéressé par un développement théorique à ce sujet pourra se reporter par exemple à la thèse de Cédric Duchène (2007). Pour étudier en pratique l'effet du jitter, j'ai calculé la transformée de Fourier discrète d'une image 320 x 240 uniforme (niveau de gris = 1 partout) échantillonnée par 200 pixels répartis selon la méthode par auto-organisation. Le résultat est présenté figure 3.16.

Le spectre du signal d'origine (une image ne comprenant qu'un seul niveau de gris) est un Dirac en zéro. On peut donc interpréter le cercle noir autour du Dirac central comme la zone fréquentielle où il n'y a pas de repliement spectral et donc pas d'artefacts.

La zone centrale n'est pas circulaire car, dans le cas d'une transformée de Fourier discrète (TFD), la valeur de la fréquence est exprimée en fraction de la fréquence d'échantillonnage. Cette fraction est de plus dépendante du nombre de points utilisés pour faire la TFD, différent ici selon la direction considérée.

En fréquence, l'extension de la zone centrale est identique dans les deux directions et vaut :

$$f_c = \frac{25}{(2 \times 320)} = \frac{20}{(2 \times 240)} = 0.041 \text{ pixels}^{-1}$$

Le gain associé à la composante aléatoire de la distribution des champs récepteurs est donc loin d'être négligeable.

### Choix de la fonction de répartition :

Il s'agit maintenant de déterminer la répartition des échantillons du champ récepteur telle que l'opération de moyenne sur ces échantillons corresponde à un filtrage passe-bas de fréquence de coupure  $f_c$  sur l'image originale.

Plaçons nous encore une fois dans le cas continu. Pour simplifier les calculs au profit du raisonnement, plaçons nous dans un espace unidimensionnel.

L'opération de moyenne sur un intervalle  $\Delta$  autour du point  $x_0$  d'une fonction  $f$  s'écrit :

$$m = \frac{1}{\Delta} \int_{x_0 - \Delta/2}^{x_0 + \Delta/2} f(x) dx$$

On peut réécrire cette équation sous la forme

$$m = \int_x f(x) \times \Pi_{\Delta}(x - x_0) dx$$

ou  $\Pi_{\Delta}$  est la fonction "Porte" normalisée définie par :

$$\Pi_{\Delta}: \mathbb{R} \rightarrow \mathbb{R}$$

$$\begin{cases} \Pi_{\Delta}(x) = \frac{1}{\Delta} \text{ pour } x \in \left[-\frac{\Delta}{2}; \frac{\Delta}{2}\right] \\ \Pi_{\Delta}(x) = 0 \text{ partout ailleurs} \end{cases}$$

Autrement dit, l'opération de moyenne est un produit de convolution entre le signal et la fonction Porte

$$m = f * \Pi_{\Delta}$$

Or on sait qu'un produit de convolution dans le domaine spatial devient un simple produit dans le domaine spectral:

$$TF(m) = TF(f) \times TF(\Pi_{\Delta})$$

L'opération de moyenne dans un domaine D autour du centre du champ récepteur s'apparente donc à un filtrage. Ce filtrage est défini par la transformée de Fourier de la fonction "Porte". La transformée de Fourier d'une fonction "Porte" est un sinus cardinal. Le module du spectre d'un sinus cardinal est présenté figure 3.17.

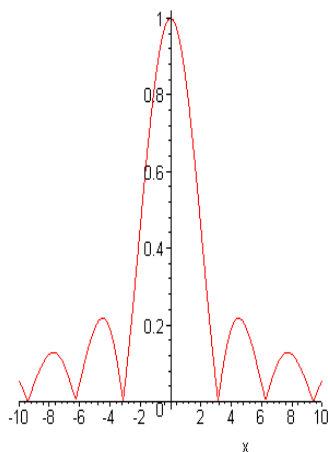


Figure 3.17: spectre en fréquence de la fonction Porte.

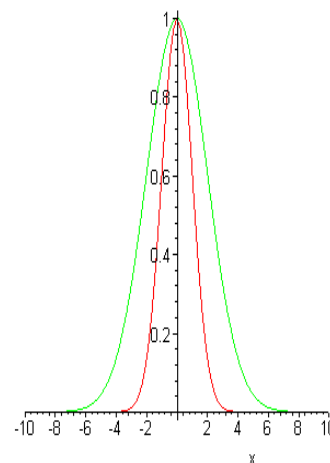


Figure 3.18 : spectre en fréquence d'un filtre Gaussien. L'extension caractéristique du filtre est défini par son écart-type  $\sigma$ : courbe rouge  $\sigma = 1$ , courbe verte  $\sigma = 2$ .

Le filtrage obtenu par une opération de moyenne est bien un filtrage passe-bas, au sens où il laisse intact le signal dans les basses fréquences (au centre). En revanche, le filtre moyennneur n'est pas un "bon" filtre passe-bas parce qu'il présente des rebonds dans les hautes fréquences et ne les supprime donc pas totalement du signal.

L'un des filtres passe-bas les plus couramment utilisés en traitement de l'image est le filtre gaussien. La figure 3.18 présente le module du spectre d'un filtre gaussien. L'intérêt de ce filtre est qu'il ne présente pas de rebond.

On sait par ailleurs que la transformée de Fourier inverse d'une gaussienne est une gaussienne.

Dans le domaine spatial, un filtrage gaussien s'obtient donc par la convolution du signal avec une gaussienne :

$$m = f * G_{\sigma}$$

ou encore :

$$m = \int_x f(x) \times G_{\sigma}(x - x_0) dx$$

Cette dernière équation peut être également interprétée comme une moyenne, mais une moyenne pondérée par une fonction Gaussienne. Par ailleurs, une moyenne pondérée peut être interprétée comme une moyenne réalisée sur des échantillons tirés aléatoirement avec une certaine densité de probabilité, en l'occurrence une densité de probabilité gaussienne.

En d'autres termes, faire une moyenne sur des échantillons répartis selon une fonction de répartition gaussienne dans un champ récepteur est comparable à opérer un filtrage gaussien de l'image. C'est donc cette répartition que nous avons choisi.

### Choix de la taille du champ récepteur :

Dans la mesure où l'on a décidé d'utiliser une fonction de répartition gaussienne, il n'y a pas, au sens strict, de bornes au champ récepteur : la fonction de Gauss est définie sur tout  $\mathbb{R}^2$ . Néanmoins, une fonction gaussienne tend vers 0 en l'infini et présente une taille caractéristique appelée écart-type, usuellement noté  $\sigma$ . La figure 3.18 présente deux gaussiennes, l'une ayant un écart-type de 1 et l'autre de 2.

La question que l'on se pose maintenant est de savoir quelle valeur d'écart-type choisir. Nous avons vu précédemment que l'on souhaitait atténuer les fréquences au delà d'une certaine valeur caractéristique, appelée fréquence de coupure et déterminée par la taille de l'image et le nombre de champs récepteurs dans l'image. Usuellement, on considère que le signal est atténué à partir du moment où le filtre réduit sa puissance de 3 décibels (le signal est alors 1000 fois plus faible).

Soit  $r$  la distance par rapport au centre du champ récepteur, la densité de probabilité des échantillons s'écrit :

$$p(r) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{r^2}{\sigma^2}}, \text{ et sa transformée de Fourier : } P(f) = e^{-2\pi^2\sigma^2 f^2}$$

Le gain du filtre associé en dB s'écrit :

$$G(f) = 20 \log_{10} \left( \frac{P(f)}{P(0)} \right) = -\frac{20}{\ln(10)} \times 2\pi^2 \sigma^2 f^2$$

A la fréquence de coupure, on veut  $G(f_c) = -3$  db, ce qui s'écrit :

$$\frac{20}{\ln(10)} \times 2\pi^2 \sigma^2 f_c^2 = 3$$

$$\text{D'où } \sigma^2 = \frac{3 \ln(10)}{20 \times 2\pi^2 f_c^2}, \text{ ou encore } \sigma = \sqrt{\frac{3 \ln(10)}{20 \times 2\pi^2}} \times \frac{1}{f_c}$$

On peut donc calculer la taille du champ récepteur permettant de vérifier ces conditions, en fonction du type de réseau considéré.

Application numérique pour une image de 320 par 240 pixels avec 200 champs récepteurs		
Réseau carré	Réseau hexagonal centré	Réseau Kohonen
5.29 pixels	4.56 pixels	3.22 pixels

On constate en particulier que la méthode d'échantillonnage par auto-organisation est la meilleure des trois puisqu'elle nécessite des champs récepteurs de petite taille : le signal est collecté au mieux grâce à l'aspect aléatoire de la position des centres des champs récepteurs.

### Choix du nombre d'échantillons :

Nous allons maintenant aborder la question du choix du nombre d'échantillons dans chaque champ récepteur. C'est cette fois-ci dans le domaine des statistiques que nous trouverons des éléments de réponse.

Le problème en l'occurrence consiste à déterminer combien d'échantillons sont nécessaire pour que la moyenne calculée sur ces échantillons (la mesure) correspondent de manière fiable à la moyenne effective de la population (l'ensemble des pixels du champs récepteur).

Connaissant :

- L'effectif de la population :  $P$
- Le nombre d'échantillons prélevés dans la population :  $n$
- La variance de la population :  $\sigma_p^2$

il est possible de déterminer la précision de la mesure (voir par exemple Melchior et Menvielle, 2009). En effet, la variance  $v^2$  de l'estimateur de la moyenne est donnée par la formule :

$$v^2 = \frac{1}{n} \left(1 - \frac{n}{P}\right) \sigma_p^2$$

On peut estimer la valeur de  $P$  à partir de l'écart-type  $\sigma_e$  choisi pour la distribution gaussienne des échantillons dans le champ récepteur. En effet, 95% des échantillons sont situés dans un cercle de rayon  $2\sigma_e$  autour du centre du champ récepteur. Le nombre total de pixels dans cette zone est donc approximativement :

$$P \sim \pi \cdot (2\sigma_e)^2$$

Application numérique :  $\sigma_e = 5$  pixels  $\Rightarrow P \sim 314$  pixels

Il est plus difficile d'obtenir une estimation de la variance de la population, c'est à dire de la valeur des pixels dans une zone de l'image de la dimension d'un champ récepteur. En effet, cette variance dépend vraisemblablement de la scène filmée, de l'optique de la caméra, des fréquences présentes dans les scènes naturelles, etc... Par ailleurs, elle dépend aussi du traitement appliqué sur l'image. Toutefois, dans le chapitre 2 nous avons vu que le filtrage et l'adaptation locale dans la rétine permettaient d'homogénéiser cette variance et en particulier qu'ils permettaient d'étaler les variations locales sur toute la plage des valeurs disponibles, autrement dit d'obtenir un histogramme local plat. Nous allons donc nous placer dans le cas où les  $P$  pixels se répartissent de manière uniforme sur tous les niveaux de gris possibles.

Soit  $L$  le nombre de niveaux de gris possibles, la variance d'une population répartie équitablement sur tous les niveaux s'écrit :

$$\sigma_p^2 = \frac{1}{L} \sum_{l=-L/2}^{L/2} l^2$$

Application numérique :  $L = 256$  niveaux  $\Rightarrow \sigma_p \sim 74$

Le graphe présenté figure 3.19 donne l'évolution de l'écart-type de la mesure de la moyenne pour un champs récepteur de taille  $\sigma_e = 5$  pixels sur une image présentant 256 niveaux de gris pour différentes valeurs de  $n$ . On constate que la courbe présente un coude pour  $n = 10$  pixels, qui correspond à la valeur optimale : en dessous, l'incertitude sur la mesure augmente très vite, en dessus, les échantillons supplémentaires ne réduisent que de peu cette incertitude.

Pour  $n=10$ , la variance calculée est d'environ 8 niveaux de gris. Nous reviendrons sur ce point dans la partie "Echelles de codage".

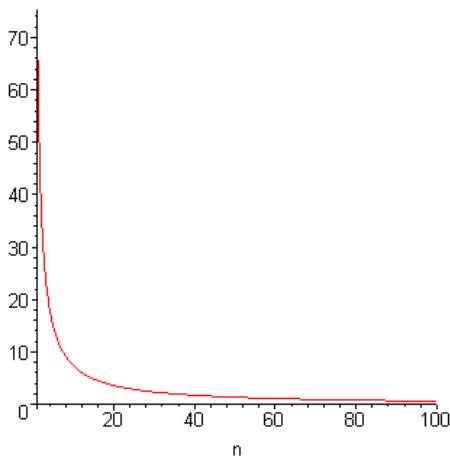


Figure 3.219: évolution de la précision de la mesure de la moyenne dans le champ récepteur en fonction du nombre d'échantillons  $n$ .

La figure 3.20 présente une rétine obtenue par la méthode d'auto-organisation de Kohonen avec les valeurs calculées dans ce paragraphe. Elle échantillonne une image  $320 \times 240$  avec 200 champs récepteurs. Chaque champ récepteur comprend 10 échantillons distribués selon une répartition gaussienne d'écart-type  $\sigma = 5$  pixels.

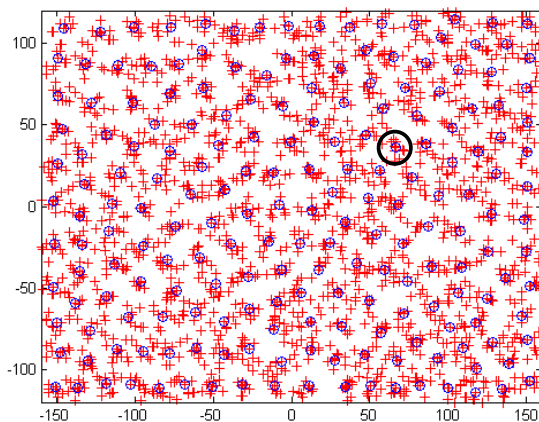


Figure 3.20 : rétine obtenue par la méthode d'auto-organisation de Kohonen. Le cercle noir indique la taille d'un champ récepteur. Grâce à l'auto-organisation de Kohonen et au théorème d'échantillonnage de Nyquist-Shannon, on a pu réduire de manière très significative la taille du champ récepteur. Ceci améliore d'autant la résolution du dispositif.

C'est ce type de rétine qui a été utilisé dans la deuxième expérience du chapitre 3 sur la localisation de cible et dans les expériences de mobilité du chapitre 5.



### 3.3.5 Echantillonnage spatialement variant

Le travail que nous avons réalisé jusqu'ici a permis de déterminer les moyens et les paramètres permettant de réaliser un échantillonnage uniforme de l'image. Toutefois, nous avons vu au chapitre 2 que la densité des photorécepteurs sur la rétine n'est pas uniforme. Elle est maximale au centre du champ visuel (au niveau de la fovéa) puis diminue selon une loi inversement proportionnelle à l'excentricité. Nous avons également vu l'intérêt que cela pouvait présenter en terme d'ergonomie du dispositif de perception.

Grâce au fichier de rétine, ce type d'échantillonnage spatialement variant est réalisable directement au niveau du dispositif, sans avoir à travailler sur l'image issue de la caméra. Pour cela, il suffit d'appliquer aux échantillons du fichier de configuration la transformation inverse de celle désirée sur l'image. La figure 3.21 présente une rétine de ce type obtenue en appliquant la réciproque de la transformation pseudo-logarithmique définie au chapitre 2. On retrouve la décroissance de la densité des champs récepteurs avec l'excentricité. La figure 3.22 illustre la projection équivalente appliquée à image.

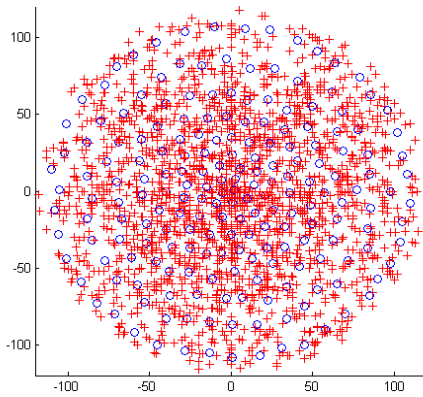


Figure 3.21 : rétine présentant un échantillonnage spatialement variant. La densité des champs récepteurs est forte au centre et décroît avec l'excentricité.



Figure 3.22 : illustration de l'effet de l'échantillonnage spatialement variant appliqué à une image. Grâce à la forte densité des échantillons, la résolution dans la zone centrale est élevée

La plupart des points discutés dans le paragraphe précédent restent valable car la transformation est appliquée après avoir placé les champs récepteurs et les échantillons de manière uniforme. En particulier, en utilisant cette méthode, la taille du champ récepteur dépend elle aussi de l'excentricité, de sorte que les propriétés obtenues concernant le repliement spectral dans le cas uniforme restent vraies dans le cas spatialement variant.

Nous reviendrons plus en détail sur ce type de rétine au chapitre 4, lorsque nous aborderons les expérimentations à ce sujet.

### 3.3.6 Correction de l'optique de la caméra

Il est également possible de corriger les déformations géométriques dues à l'optique de la caméra, en particulier dans le cas des caméras grands angles qui produisent des déformations importantes (fig. 3.23). Ceci s'avère nécessaire dans notre cas puisque nous voulons étudier l'impact de la projection de l'image et devons donc être à même de bien la contrôler.

Le principe est similaire à celui de l'échantillonnage spatialement variant. La déformation de

l'optique de la caméra est calculée grâce à un logiciel (Matlab Calibration Toolbox) disponible sur Internet. Cette fois, on applique directement cette déformation (et non la déformation inverse) aux échantillons du fichier de configuration (fig. 3.24). On annule par ce procédé la déformation induite par l'optique de la caméra.

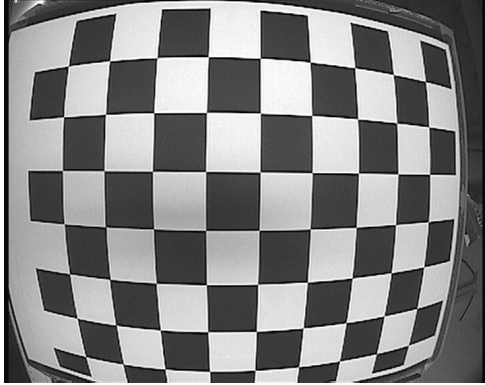


Figure 3.23 : photo d'un échiquier avec la caméra grand angle CWHQX. La déformation optique produit un effet de loop bien connu.

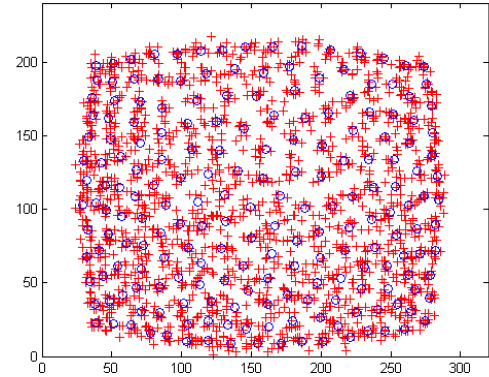


Illustration 3.24 : Rétine "uniforme" avec correction de la déformation optique de la caméra. En reproduisant la déformation de la caméra sur les échantillons du fichier de configuration, on rétabli une

Une documentation complète du logiciel et de ses paramètres est disponible sur Internet (Bouquet, 2008), ainsi que les fonctions qui permettent d'appliquer une déformation identique à celle de la caméra sur une image.

Cette correction peut bien sûr être composée avec une autre projection, comme par exemple celle définie dans le paragraphe précédent.

### 3.3.7 Conclusion

Dans cette partie, nous avons mis en place une méthode de sous-échantillonnage de l'image basée sur une répartition par auto-organisation de champs récepteurs dans l'image et par une distribution gaussienne des échantillons dans chaque champ récepteur dont la taille est calculée en relation avec le théorème d'échantillonnage de Shannon et la limite de Nyquist.

Une question reste toutefois entière à ce sujet : doit-on ou non respecter le théorème de sous-échantillonnage de Shannon? Les résultats obtenus à l'aide de systèmes de réalité virtuelle simulant l'échantillonnage dans une prothèse visuelle semble en effet montrer que les sujets restent performants y compris au delà de la limite de Nyquist (Chen *et al.*, 2004). Selon les auteurs, cela pourrait être expliqué par une "intégration des mouvements" de l'utilisateur. Dans une étude ultérieure, les auteurs ont étudiés plus précisément l'effet du mouvement sur les performances des sujets.

Cette référence au mouvement laisse à penser qu'une étude de cette question de fréquence limite dans un cadre sensori-moteur pourrait s'avérer particulièrement riche. Nous avons en tout cas posé ici les bases pour étudier cette question : l'utilisateur est libre de choisir de ne pas respecter cette limite, il lui suffit pour cela de paramétrer la taille des champs récepteurs à une valeur inférieure à celles calculées ici.

## 3.4 Codage sonore

Nous allons maintenant aborder en détail la question de la transcription en son des valeurs issues de l'opération d'échantillonnage présentée dans la section 2. Avant d'aborder en pratique cette question, je vous propose quelques éléments de base sur le fonctionnement du système auditif humain.

### 3.4.1 Le système auditif humain

Le système auditif est le système sensoriel stimulé par les variations de pression de l'air. Chez l'homme, ce dispositif sensoriel se situe au niveau des oreilles. Les variations de pression acoustique mettent en mouvement une membrane : le tympan. Ce mouvement est transmis à l'entrée de la cochlée via deux petits osselets appelés marteau et enclume (fig. 3.25).

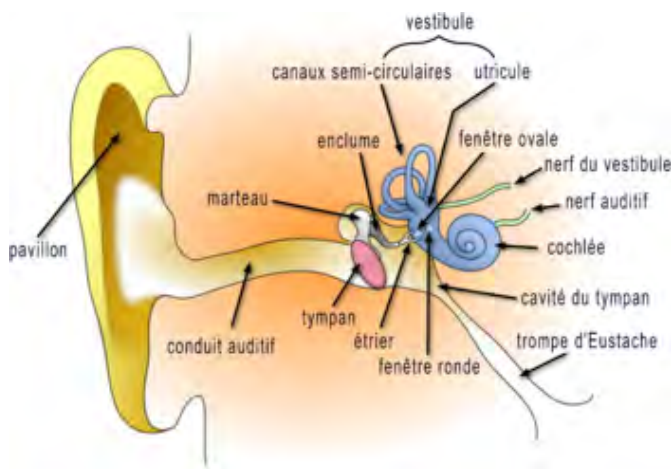


Figure 3.25 : le système auditif humain.

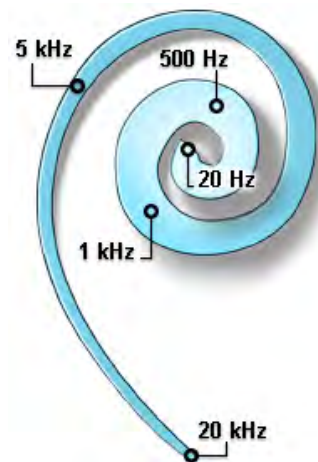


Figure 3.26 : la membrane basilaire Grâce à la géométrie de la cochlée, le signal mécanique se décompose selon ses fréquences.

La cochlée est l'organe où le signal mécanique est converti en signal électrique puis propagé vers le cerveau par le nerf auditif. Il s'agit d'une sorte d'appendice enroulé sur lui-même et rempli d'un liquide appelé endolymphe. Sa structure géométrique permet de décomposer le signal mécanique selon ses composantes fréquentielles. En effet, la fréquence de résonance mécanique du système n'est pas la même selon la position que l'on considère dans la cochlée. Ainsi, à l'entrée de la cochlée, la fréquence de résonance est de l'ordre de 20 Hz, et augmente jusqu'à 20 kHz à son extrémité.

La cochlée est par ailleurs séparée en deux dans le sens de la longueur par la membrane basilaire (fig. 3.26). Cette dernière est tapissée de cellules ciliées qui convertissent le signal mécanique en influx nerveux. Le signal acoustique est ainsi décomposé dans le domaine fréquentiel avant d'être converti en signal nerveux. Ce mécanisme est vraisemblablement à la base de la capacité de l'humain à identifier les différentes fréquences qui composent un signal acoustique naturel.

Le système auditif humain n'est pas uniquement sensible aux différentes composantes fréquentielles d'un signal. Il est également sensible à leur intensité respective. Cette dernière est en effet codée par les cellules ciliées, dont le taux de décharge dépend de l'intensité de la stimulation mécanique qu'elles subissent. De cette manière, le sujet humain est capable de sentir les variations d'intensité respective des différentes composantes fréquentielles du signal. C'est en grande partie sur

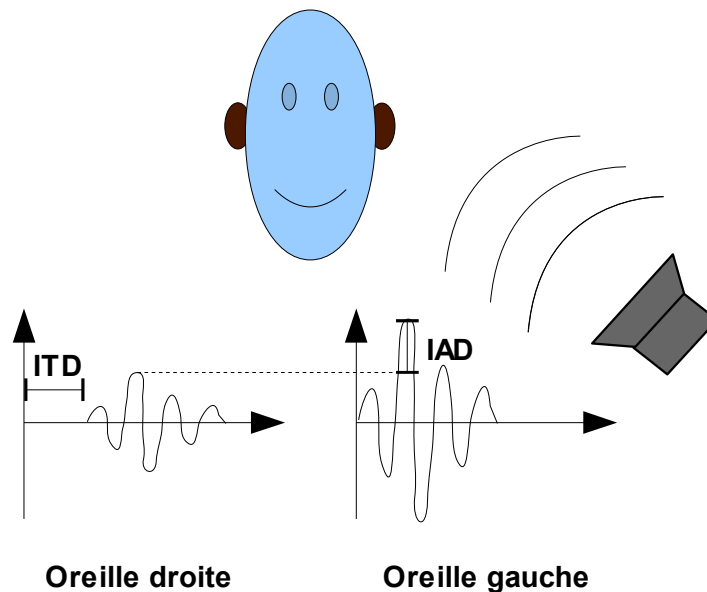


Figure 3.27 : les indices binauraux. Si la source n'est pas en face du sujet, des différences apparaissent entre le signal reçu par l'oreille gauche et celui reçu par l'oreille droite : une différence d'amplitude (ILD) et une différence de phase (ITD).

cette capacité que repose la compréhension de la parole chez l'homme, et l'esthétique musicale.

Une dernière propriété du système auditif humain enfin est la capacité de distinguer les signaux acoustiques selon la localisation spatiale de leur source. Cette faculté provient notamment de l'existence de différences entre le signal arrivant à l'oreille droite et celui arrivant à l'oreille gauche, en fonction de la position de la source.

On isole en particulier deux paramètres qui dépendent de l'azimut de la source, c'est à dire de sa position dans le plan horizontal (Rayleigh, 1907) :

1. La Différence Interaurale de niveau sonore (ILD pour Interaurale Loudness Difference). Selon sa position, le signal généré par la source va être plus ou moins masqué par la tête, la stimulation sonore au niveau de l'oreille du côté de la source est plus forte que dans l'oreille opposée.
2. La Différence Interaurale de Temps (ITD pour Interaurale Time Difference). En fonction de la position de la source, le trajet du signal jusqu'à chaque oreille est légèrement différent. Cette différence va générer une différence de phase entre les deux ondes. Si le système auditif n'est pas sensible à la phase globale du signal, il est en revanche sensible à une différence de phase entre les signaux de chaque oreille.

Ces deux paramètres sont illustrés figure 3.27.

ILD et ITD sont des paramètres couramment utilisés pour faire ce que l'on appelle de la synthèse binaurale, c'est à dire pour simuler le son produit par une ou plusieurs sources à différentes positions horizontales devant le sujet. Dans le cas de TheVIBE, c'est ce type de synthèse qui est utilisé.

Les capacités de localisation du système auditif humain ne se limitent pas à la discrimination de la position azimutale de la source. La discrimination de l'élévation de la source, par exemple, est

également possible. C'est toutefois un problème plus complexe qui fait intervenir entre autre le filtrage par le pavillon de l'oreille. Le lecteur intéressé pourra se référer à Blauert, 1997.

### 3.4.2 Génération du signal sonore dans TheVIBE

Le fonctionnement du dispositif TheVIBE repose sur la capacité du système auditif à discerner les différentes sources formant un signal sonore selon leur fréquence et selon leur azimuth, de la même manière que le système visuel discerne différentes sources de lumières dans une image selon leur position horizontale et leur position verticale. L'intensité de chaque "source lumineuse" est par ailleurs transcrite par l'intensité respective de chaque source sonore.

En pratique, le signal généré par TheVIBE consiste en deux signaux ( $S_l$  à gauche et  $S_r$  à droite) formés par sommation de  $N$  signaux élémentaires sinusoïdaux,  $N$  étant le nombre de champs récepteurs distribués dans l'image. Ces  $N$  signaux élémentaires correspondent aux  $N$  sources sonores qui forment le signal

Chaque signal élémentaire est caractérisé par :

1. Une fréquence :  $\omega_i$
2. Une phase globale :  $\Phi_i$
3. Une amplitude globale :  $A_i$
4. Deux coefficients de panoramique  $P_i^l$  et  $P_i^r$ , différents selon l'oreille gauche ou droite et déterminés conjointement par un paramètre d'azimut  $a_i$  compris entre 0 (à gauche) et 1 (à droite).
5. Une différence de temps  $\tau_i$  au niveau du signal droit seulement

$$\begin{cases} S_l = \sum_{i=1}^N A_i \times P_i^l \times \sin(\omega_i \times t + \phi_i) \\ S_r = \sum_{i=1}^N A_i \times P_i^r \times \sin(\omega_i \times (t + \tau_i) + \phi_i) \end{cases}$$

Ces signaux sont convertis par la carte audio du PC en signaux électriques, eux-mêmes convertis en pression acoustique. Le son généré par TheVIBE est donc une somme de sons purs de différentes intensités.

A l'exception des valeurs d'amplitude  $A_i$  déterminées par le logiciel pour chaque image, les paramètres de chaque sources sont fixés par l'utilisateur dans le fichier de configuration. Leur rôle est de faire en sorte que le système auditif soit capable de les différencier les unes des autres, à partir de leur caractéristique fréquentielle et de leur position apparente dans le plan azimuthal, y compris lorsqu'elles "fonctionnent" en même temps.

Pour choisir ces paramètres, nous allons nous intéresser en détail à la relation que ces paramètres entretiennent avec les caractéristiques sensorielles décrites précédemment. Pour cela, nous allons aborder le domaine de la psychophysique.

### 3.4.3 Choix des paramètres des signaux élémentaires

Le "son pur" est l'évènement sensoriel produit par la stimulation du système auditif humain par un signal acoustique de forme sinusoïdale. Ce terme est issue de la psychophysique.

La psychophysique est sans doute le courant le plus marquant consacré à l'étude de la sensation. Il a été initié par les travaux de Fechner en 1850 et s'est attaché depuis à étudier "les relations quantitatives démontrées entre des événements physiques identifiés et mesurables et des réponses évoquées selon une règle expérimentale avérée" (Bonnet, 1986, p9).

Comme son nom le laisse supposer, la psychophysique n'a pas eu pour seule ambition d'étudier les relations entre physique et psychologie. Elle s'est aussi attachée à établir une véritable "physique" de la sensation, c'est à dire à mettre en place un cadre mathématique permettant de décrire la nature et les lois de cet "espace sensoriel" : "la psychophysique s'est développée [...] pour représenter les événements psychologiques et *leurs relations* au moyens de nombres dont les relations (mathématiques) correspondent à celles qu'entretiennent les événements psychologiques mesurés" (Bonnet, 1986, p126). La notion même d'espace sensoriel est typiquement issue de la psychophysique. Cette dernière a en effet postulé l'existence de *dimensions* selon lesquelles la sensation provoquée par une stimulation peut être caractérisée.

Dans le cas du son pur, on distingue essentiellement trois dimensions : la sonie, la tonie, et le panoramique. La sonie correspond à ce que l'on nomme usuellement l'intensité, ou le volume sonore. La tonie correspond à la notion de hauteur tonale, et l'azimut, correspond à la position horizontale apparente d'une source sonore (réelle ou virtuelle) d'où serait issue le son pur. On reconnaît bien sûr les notions abordées précédemment.

Le lien, établi dans le cadre de la psychophysique, entre paramètres physiques et dimensions sensorielles prend la forme de fonctions dites psychométriques. C'est à l'aide de ces fonctions que nous allons travailler sur le codage sonore dans TheVIBE.

### 3.4.4 Codage de l'intensité sonore

Dans TheVIBE, seul le paramètre d'amplitude globale de chaque source sonore ( $A_i$ ) est calculé pour chaque image. Il permet de faire le lien entre l'activité calculée dans le champ récepteur  $i$  (la moyenne sur les échantillons du champ récepteur  $i$ , que nous noterons  $Act_i$ ) et l'intensité sonore de la source  $S_i$  associée.

Dans la version initiale du logiciel, l'amplitude de chaque signal élémentaire était proportionnelle à l'activité du champ récepteur associé :

$$A_i = \alpha Act_i$$

Ce type de codage a été utilisé dans la première expérience de localisation (ch. 4).

Cette relation linéaire entre l'activité du champ récepteur et l'amplitude du signal n'est néanmoins pas vraiment adaptée au fonctionnement du système auditif. Sur le plan physique tout d'abord, l'intensité (ou la puissance) du signal sonore n'est pas liée à l'amplitude du signal sonore mais à son carré. De plus, en psychophysique, on travaillera non pas directement sur la puissance du signal, mais sur le logarithme de cette puissance. L'échelle usuelle pour représenter l'intensité du signal sonore en psychophysique est en effet le décibel définie par :

$$L(dB) = 10 \times \log_{10}(A_i^2) = 20 \times \log_{10}(A_i)$$

L'utilisation d'une fonction logarithmique dérive des travaux de Weber et Fechner sur les capacités de discrimination des systèmes sensoriels, plus précisément sur la notion de seuil différentiel juste perceptible (Just Noticeable Difference en anglais, noté JND). Pour une valeur donnée de l'intensité d'un signal sonore, le JND correspond à la différence minimale d'intensité au dessous de laquelle le sujet ne parvient pas à distinguer deux stimulations.

Les mesures réalisées par Weber le conduisent à postuler que, pour un type de stimulation donné, le seuil différentiel est proportionnel à la grandeur physique. Par exemple, pour discriminer selon leur hauteur tonale, deux sons autour de 1000 Hz, le seuil différentiel sera deux fois plus élevé que le seuil différentiel à 500 Hz. En notant  $\Delta I$  ce seuil différentiel, on peut écrire :

$$\frac{\Delta I}{I} = k$$

Cette observation a conduit Fechner à définir l'une des premières fonction psychométriques en proposant d'utiliser le logarithme de la grandeur physique :

$$\Psi(I) = \ln(I)$$

L'intérêt de cette proposition réside dans le fait que sur cette échelle psychométrique, le seuil différentiel est constant. En effet, soit  $\Delta F$  le seuil différentiel pour une valeur  $F$  de cette échelle, on peut écrire :

$$\Delta F = \Psi(I) - \Psi(I + \Delta I)$$

où  $\Delta I$  est le seuil différentiel exprimé dans l'échelle physique. En considérant par ailleurs  $\Delta I \ll I$ , on peut considérer :

$$\Delta F = \Psi'(I) \Delta I = \frac{\Delta I}{I} = \text{constante}$$

L'utilisation du logarithme décimal plutôt que du logarithme naturel et de l'échelle en décibels est issue des usages en traitement du signal. Cela ne change en rien les propriétés de la fonction psychométrique de Fechner.

### **Intérêt dans le cas du dispositif TheVIBE**

D'autres fonctions psychométriques ont été construites après les travaux de Fechner. Stevens en particulier a proposé une nouvelle méthode de mesure psychophysique où le sujet doit quantifier le ratio entre deux signaux. Cette méthode a aboutie à la création d'une nouvelle unité : le sone. Stevens a par ailleurs proposé de modéliser les résultats obtenus en sonnes par une loi de puissance. Cette loi conserve la notion de ratio, c'est à dire que si on considère 4 stimulus d'intensité  $I_1$ ,  $I_2$ ,  $I_3$  et  $I_4$  vérifiant

$$\frac{I_1}{I_2} = \frac{I_3}{I_4} \quad \text{alors} \quad \frac{\Psi(I_1)}{\Psi(I_2)} = \frac{\Psi(I_3)}{\Psi(I_4)}$$

La question de savoir laquelle des deux fonctions, de Fechner ou de Stevens, représente mieux la réalité telle qu'elle est formalisée dans le cadre de la psychophysique est toujours sujet à controverse (Bonnet, 1986, p194).

Dans le cas présent, la propriété de la fonction de Fechner m'a paru particulièrement pertinente au regard de l'approche sensori-motrice de la perception. En effet, le seuil différentiel, exprimé dans l'échelle de Fechner est constant. Cela signifie qu'une variation de la stimulation correspondant à un certain  $\Delta \Psi$  a la même probabilité d'être détectée quelque soit la valeur de  $\Psi$ . On privilégie donc ici la capacité du système sensitif à détecter les variations du signal sonore, plutôt qu'une métrique quelconque des valeurs associées à la stimulation. Par ailleurs, Garner et Hake (1951) expliquent que l'utilisation d'une fonction égalisant les capacités de discrimination du système sensitif maximise sa capacité de transfert, au sens de la théorie de l'information. Nous reviendrons sur cette question par la suite.

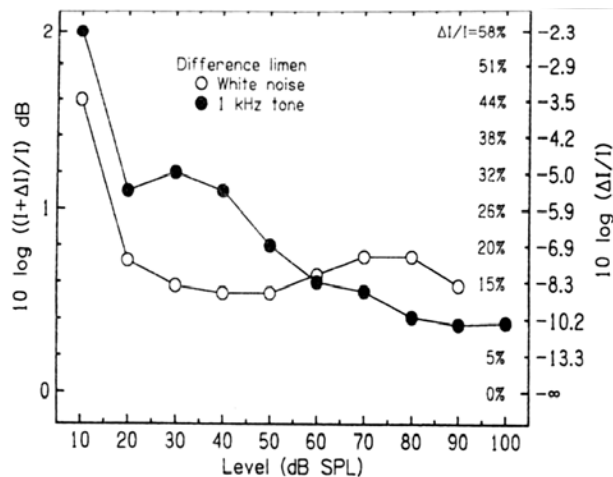


Figure 3.28 : mesure de la différence juste notable d'intensité sonore pour un son pur à 1000 Hz et pour un bruit blanc (Houtsma, 1980 reproduit par Hartmann, 1997)

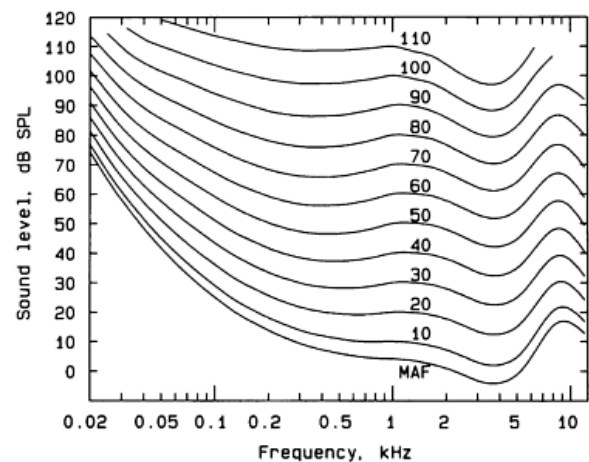


Figure 3.29 : les indices binauraux. Pour une large plage de fréquences, les variations du signal génèrent de variations de sonie équivalentes.

### Validité de l'hypothèse de Weber pour l'échelle d'intensité sonore en dB

La figure 3.28 présente les mesures de JND pour un son pur à 1000 Hz et pour un bruit blanc, pour des intensités sonores variant entre 10 et 100 dB SPL. La mesure en dB SPL est une échelle en dB ramenée au seuil de perception auditif humain (qui correspond au 0 de l'échelle).

Dans le cas de TheVIBE, on utilise des sons purs compris entre 30 et 90 dB, ce qui correspond au domaine de confort de l'oreille. Dans ce domaine d'intensité, le JND relatif pour un son pur à 1000 Hz est relativement stable et compris entre 10 et 30%. L'utilisation de l'échelle d'intensité en dB est donc convenable. Une sophistication ultérieure à partir de mesures psychophysiques pour obtenir une meilleure égalisation du JND relatif reste toutefois envisageable (Garner, 1953).

### Interaction entre la sonie et la fréquence

Il existe un troisième méthode de mesure en psychophysique : la méthode des ajustements. Le sujet doit, pour ce type de mesure, ajuster un paramètre physique du signal de manière à égaliser sa valeurs psychophysique avec celle d'un stimulus de référence. Ici, le stimulus de référence est une sinusoïde à 1000 Hz, avec différents niveaux d'intensité sonore. L'unité de mesure ainsi obtenue a été appelé le "phone". La figure 3.29 représente des courbes "d'isophonies".

Ces courbes montrent que la sonie d'un son pur dépend non seulement de son niveau exprimé en dB mais également de la fréquence du son pur. Cette dépendance est limitée dans la gamme de fréquence comprise entre 300 et 3000 Hz, mais reste néanmoins sensible, avec des variations allant jusqu'à 10 dB de différence entre deux signaux, pour une même intensité du signal (en dB).

On remarquera néanmoins que, dans cette même gamme de fréquences, l'écart entre chaque courbe d'isophonie est relativement constant (et de fait égal à l'écart d'intensité sonore en dB sur le stimulus de référence, i.e. 10 dB). Ainsi, si le niveau de sonie varie, pour une même intensité du signal, en fonction de la fréquence du signal, les variations de sonie associées à une variation d'intensité du signal sont, elles, relativement indépendante de la fréquence. Cette remarque rejoint de nouveau le cadre théorique dans lequel on s'inscrit. Dans une approche sensorimotrice, ce sont les variations des valeurs et non les valeurs elles même qui ont un sens. A ce titre, l'interaction



observée entre la sonie et la fréquence du stimulus dans la gamme 300-3000 Hz est peu susceptible de réduire la qualité du couplage sensorimoteur.

Sous cette hypothèse, l'utilisation d'une échelle d'intensité sonore en dB est donc toujours admissible. Pour toutes les expériences postérieures à la première expérience sur la localisation, c'est cette échelle que nous avons utilisée.

#### En pratique

L'activité moyenne des champs récepteurs est calculée par

$$Act_i = \frac{1}{n} \sum_{k=1}^n l_{i,k}$$

où  $l_{i,k}$  est le niveau de gris (compris entre 0 et 255) du  $k^{ième}$  pixel échantillonné dans le récepteur  $i$ .

Dans la version d'origine de TheVIBE, utilisée pour la première expérience de localisation, l'amplitude de la source du champs récepteur est proportionnelle à l'activité et comprise entre 0 et 1:

$$A_i = \frac{1}{255} \times Act_i$$

Dans les versions suivantes, le codage de l'amplitude est fait en dB, sur une plage totale de 60 dB.

$$A_i = 10^{(-3)} \times 10^{\frac{3 Act_i}{255}}$$

### 3.4.5 Codage du panoramique de la source

Nous allons maintenant nous intéresser aux paramètres du fichier de configuration permettant de manipuler la position azimuthale apparente de la source. Ces paramètres sont l'ILD et l'ITD définis précédemment.

#### **Lien entre position azimuthale et ILD**

L'ILD est le paramètre le plus couramment utilisé pour simuler la position azimuthale d'une source sonore. Il est utilisé sur les dispositifs stéréo usuels comme par exemple les autoradio ou les chaînes stéréo où l'on peut régler la "balance" entre le haut parleur gauche et le haut parleur droit.

L'ILD correspond à la différence de niveau du signal entre les deux oreilles exprimée en dB. Les valeurs d'azimut extrêmes (stimulus complètement à gauche ou complètement à droite) sont atteintes pour des ILD de l'ordre de 15 à 20 dB (Kietz, 1953, cité dans Blauert, 1997). Néanmoins, au delà d'une dizaine de dB, l'imprécision sur l'azimut devient importante.

Entre ces valeurs extrêmes, l'azimut apparent dépend à peu près linéairement de l'ILD (fig. 3.30). Par ailleurs, le JND de l'ILD est à peu près constant dans ce domaine (Herchkowitz et Durlach, 1969, Hafter *et al.*, 1977) et de l'ordre de 1 dB. Une échelle linéaire de l'ILD en dB correspond donc à une échelle d'égale discrimination, dont l'intérêt dans le cadre d'une approche sensorimotrice a déjà été discuté.

Voyons maintenant l'interaction entre la position apparente d'un son pur pour une ILD donnée en

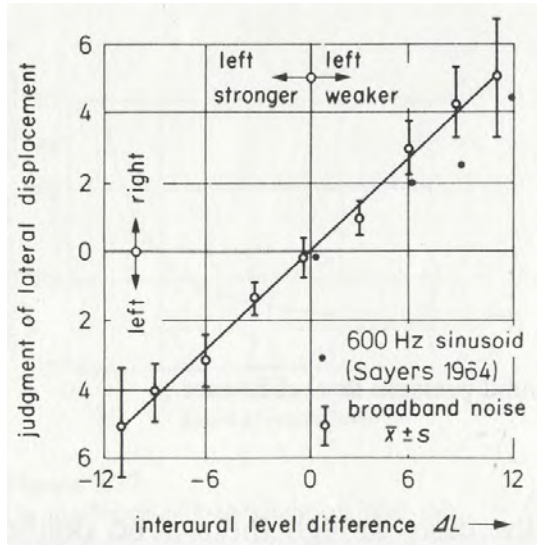


Figure 3.30: lien entre la différence d'amplitude interaurale en dB et l'azimut apparent de la source sonore pour un son pur et un bruit à large bande (d'après Blauert, 1997).

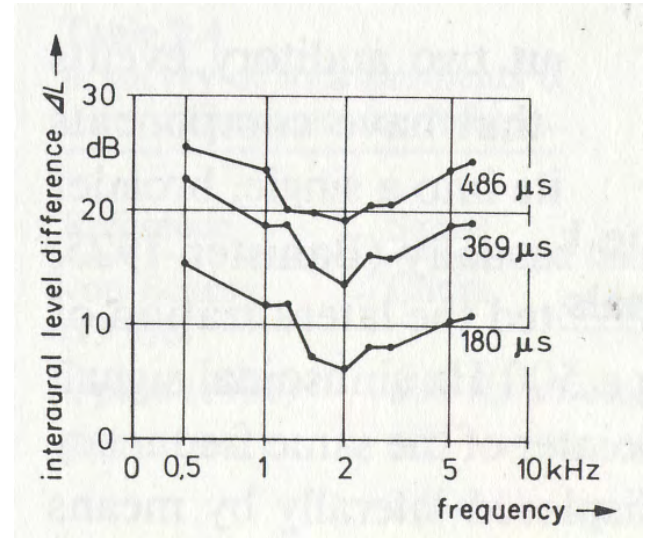


Figure 3.31 : mesure d'ajustement de l'IAD en dB pour trois positions de référence en fonction de la fréquence. (Feddersen, 1957, d'après Blauert, 1997). Les valeurs en microsecondes indiquent l'ITD du pointeur de référence.

fonction de sa fréquence. La figure 3.31 est issue des travaux de Feddersen qui a demandé à des sujets d'ajuster la position apparente d'un son pur avec celle d'un bruit blanc présentant une ITD donnée constituant une position apparente de référence. On constate sur cette figure que l'azimut d'un signal pour une ILD donnée dépend de la fréquence du son pur (presque du simple au double), y compris dans la gamme de fréquence 300-3000Hz.

Néanmoins, tout comme dans le cas de la sonie, les variations de l'ILD en fonction de la fréquence semblent dépendre peu de l'azimut du stimulus de référence. Encore une fois, si l'azimut du stimulus dépend de la fréquence, les variations de l'azimut du stimulus n'en dépendent presque pas. On peut donc à nouveau insister sur le fait que la notion de couplage sensori-moteur privilégie les variations de la stimulation et non la valeur associée à la stimulation en elle même.

Contrairement au cas de la sonie néanmoins, l'azimut est, dans le cas du dispositif TheVIBE, voué à être utilisé conjointement avec la tonie du son pur pour émuler l'espace bidimensionnel dans lequel sont disposés les champs récepteurs. Il se peut alors que, du fait des interactions observées entre azimut et tonie, les variations de la stimulation lors du passage d'un champ récepteur à un autre soient particulièrement irrégulières (transformations non séparables). Ceci aurait pour conséquence possible de rendre l'apprentissage des lois de contingences sensori-motrices associées à la topologie 2D de l'image difficile, voir impossible. Dans ce cas, un codage du panoramique qui dépende de la fréquence de la source est envisageable.

#### Lien entre position azimutale et ITD :

Le deuxième élément sur lequel se base le système auditif pour séparer les sources sonores en fonction de leur azimut est l'Interaural Time Difference, qui correspond à la différence de marche entre le signal arrivant à l'oreille gauche et celui arrivant à l'oreille droite.

Cet indice est toutefois moins intéressant dans notre cas que l'ILD. En effet la différence de temps joue sur la différence de phase entre les deux signaux. A partir d'une certaine fréquence, cette différence de phase devient de fait ambiguë, car la différence de marche peut être de plusieurs

cycles. En pratique, pour des sons purs, le système auditif n'est plus sensible à l'ITD au delà de 1500 Hz. Par ailleurs la question du lien entre ITD et ILD est assez délicate et dépend en particulier de la bande de fréquence occupée par le signal (voire Blauert, 1997, p 164 et p 401).

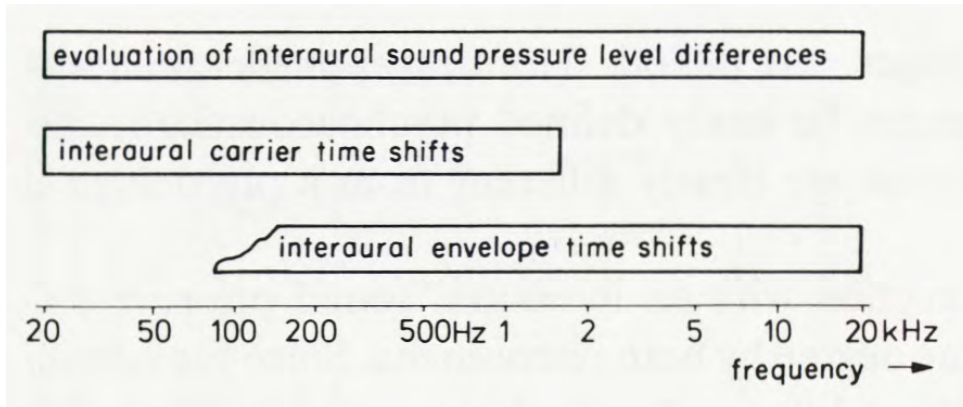


Figure 3.32: Domaine d'utilisation des indices inter-auraux.

L'utilisation de l'ITD dans TheVIBE pourrait en fait s'avérer intéressante à condition de travailler non plus avec des sons purs mais sur des sons à large bande. Dans ce cas, le système auditif est capable de discriminer la différence interaurale de temps sur l'enveloppe du signal. Il n'y a plus alors de problème d'ambiguïté, cet indice est valable dans presque toute la gamme de fréquences à l'exception des fréquences inférieures à 100 Hz (fig 3.32).

Si l'utilisation de sons large bande n'était pas envisageable à court terme parce que cela aurait nécessité une réécriture complète du programme, il m'a semblé intéressant de creuser une question au sujet de l'ITD : l'ITD peut-elle permettre d'empêcher la fusion de deux sons purs de même fréquence, c'est à dire empêcher qu'ils soient perçus comme un seul et même son pur ?

Pour répondre à cette question on considère deux sources sonores S1 et S2 dans TheVIBE, de même fréquence  $\omega$  mais d'amplitude, d'IAD et d'ITD différentes, la somme de ces deux sources est-elle équivalente à une seule source sonore ?

Pour simplifier les calculs, on considérera une source S2 de référence, d'amplitude 1, d'IAD, de phase globale et d'ITD nulles. On utilisera par ailleurs la représentation complexe. Le signal gauche sera signalé par un l en exposant, le signal droit par un r. Le signal stéréo des sources S1 et S2 s'écrit donc :

$$\begin{cases} S_1^l = A P^l e^{i(\omega t + \phi)} \\ S_1^r = A P^r e^{i(\omega(t + \tau) + \phi)} \end{cases} \quad \begin{cases} S_2^l = e^{i\omega t} \\ S_2^r = e^{i\omega t} \end{cases}$$

Calculons maintenant la somme algébrique de ces deux signaux telle qu'elle serait effectuée dans TheVIBE :

$$\begin{cases} S_1^l + S_2^l = A P^l e^{i(\omega t + \phi)} + e^{i\omega t} \\ S_1^r + S_2^r = A P^r e^{i(\omega(t + \tau) + \phi)} + e^{i\omega t} \end{cases}$$

En mettant l'exponentielle en facteur :

$$\begin{cases} S^l = S_1^l + S_2^l = e^{i\omega t} (A P^l e^{i\phi} + 1) \\ S^r = S_1^r + S_2^r = e^{i\omega t} (A P^r e^{i(\omega\tau + \phi)} + 1) \end{cases}$$

Le signal S est donc toujours un son pur de fréquence  $\omega$ . La question de son amplitude et de sa phase globale n'est ici d'aucun intérêt puisque ces deux paramètres n'interviennent pas dans la position apparente de la source. Ce sont les paramètres interauraux (ILD et ITD) du signal résultant S qui vont déterminer en revanche si les signaux  $S_l$  et  $S_r$  correspondent ou non à une source sonore d'azimut donné. Ils sont exprimés par le module et l'argument du rapport  $S_l / S_r$ . Celui-ci s'écrit :

$$\frac{S_l}{S_r} = \frac{A P^l e^{i\phi} + 1}{A P^r e^{i(\omega\tau + \phi)} + 1}$$

Quoiqu'il advienne, ce nombre complexe admet un module et une phase, ce qui signifie que le signal S peut s'exprimer sous la forme "générique" d'une seule source sonore dans TheVIBE. Néanmoins, ces valeurs d'ILD et d'ITD peuvent ne pas être cohérentes. Pour le savoir, la question du lien entre ILD et ITD devra être résolue.

Cela dit, dans le cas où l'on n'utilise pas le paramètre d'ITD ( $\tau = 0$ ), on remarquera que la phase globale choisie pour chaque source sonore provoque l'apparition d'une ITD dans le signal S : sans cette phase, le quotient  $S_l / S_r$  serait réel. Ainsi, si l'on n'utilise pas le paramètre ITD dans la génération des signaux, la somme de deux signaux sonores de même fréquence ne donne pas un signal admissible, au sens mathématique du terme, dans TheVIBE. Il n'est pas garantie toutefois que cela permette au système auditif de discriminer s'il s'agit de deux sources différentes ou non.

Dans la pratique, nous n'avons donc pas utilisé l'indice ITD.

#### En pratique :

La première expérience réalisée avec la première version du logiciel utilisait le codage par défaut du panoramique. Dans ce dernier l'amplitude des coefficients de panoramique dépendait linéairement du paramètre  $a_i$ , lu dans le fichier de configuration. On avait alors :

$$\begin{cases} P_i^l = a_i \\ P_i^r = 1 - a_i \end{cases}$$

Dans les autres expériences, les coefficients de panoramique sont calculés de manière à ce que l'ILD du signal dépende linéairement du paramètre d'azimut  $a_i$ . La gamme d'ILD possible s'étend de -12 dB à +12dB avec :

$$\begin{cases} P_i^l = \frac{1}{10^{\left(\frac{6}{20}\right)}} \times 10^{-6 \times \frac{(2a_i - 1)}{20}} \\ P_i^r = \frac{1}{10^{\left(\frac{6}{20}\right)}} \times 10^{6 \times \frac{(2a_i - 1)}{20}} \end{cases}$$

### 3.4.6 Codage de la tonie

La tonie est définie, dans le standard ANSI (standard Américain), par "cet attribut de la sensation auditive selon lequel les sons peuvent être classés sur une échelle de valeurs allant du grave à l'aigu. La tonie dépend principalement du contenu fréquentiel du stimulus sonore, mais aussi de son intensité et de la forme d'onde du stimulus"<sup>5</sup>.

5 "Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends mainly on the frequency content of the sound stimulus, but it also depends on the sound

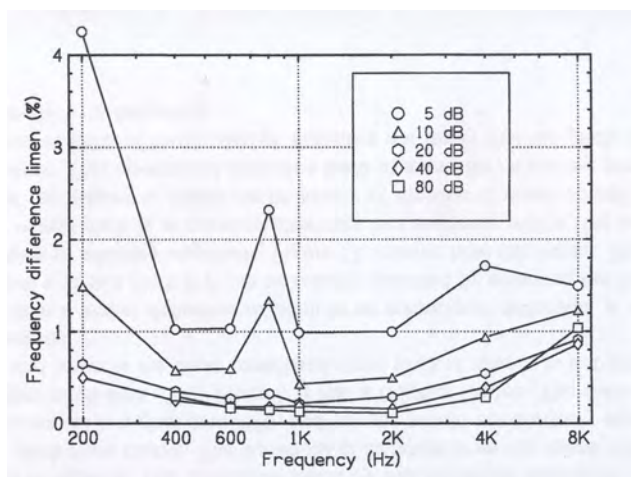


Figure 3.33 : Just Noticeable Difference en fréquence pour des sons purs de différentes intensités. Wiek *et al.*, 1977, d'après Hartmann, 1997.

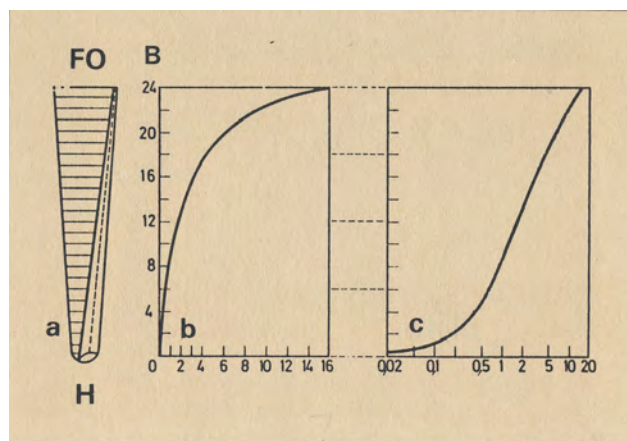


Figure 3.34 : bandes critiques et échelles en Bark. En a, alignement des bandes critiques le long de la membrane basilaire ; en b et c, correspondance entre la bande critique (en ordonnée) et la fréquence en abscisse (échelle linéaire en b et logarithmique en c), d'après Buser et Imbert, 1987.

Tout comme pour la sonie, plusieurs échelles psychophysiques de tonie ont été définies, selon le déterminant comportemental considéré. Par ailleurs, la capacité du système auditif à discriminer deux fréquences est essentiellement proportionnelle à la fréquence du signal. Dans une large gamme l'hypothèse de Weber est vérifiée (fig. 3.33), et une échelle logarithmique en fréquence est donc convenable pour le codage de la tonie dans le logiciel TheVIBE (voir discussion au §3.3.1).

La tonie dépend également de la sonie, comme le souligne le standard ANSI. Pour une augmentation de 30 dB, les courbes d'isotonie présente des variations de l'ordre de 10-15 Hz pour toutes les fréquences (Klein, 1981, cité dans Blauert, 1997).

Une distinction importante entre la sonie et la tonie a été soulignée par Stevens. Ce dernier propose de distinguer, en psychophysique, les continuum "prothétiques" et les continuum "métathétiques". Les continuum prothétiques correspondent à des processus physiologiques additifs comme la sonie : augmenter l'intensité d'un signal sonore augmente la fréquence des potentiels d'actions émis par les neurones sélectionnés. Ainsi, un son ne peut pas posséder deux sonies, les intensités des stimulus se cumulent.

La tonie en revanche est un continuum métathétique, dans le sens où la variation de la fréquence du signal entraîne l'activation d'un ensemble différent de neurones. En ce sens, le processus physiologique mis en jeu n'est plus additif. Plusieurs fréquences peuvent alors être discriminées dans un seul son.

Dans le cas des continuum métathétiques, un nouvel élément intervient dans la capacité de discrimination du système sensoriel : le masquage. Dans un signal constitué de deux fréquences suffisamment proches, la fréquence la plus énergétique peut masquer l'autre composante du signal. Le sujet ne perçoit alors plus que la composante de forte énergie, c'est à dire un son pur.

Les expériences de masquage ont conduit à la détermination de bandes critiques, qui délimitent des intervalles de fréquence dans lesquels on observe un phénomène de masquage. Si deux

pressure and the waveform of the stimulus" (American National Standard (ANSI), 1994, from Hartmann, 1997.

fréquences sont situées dans deux bandes critiques différentes, leur interaction ne sera pas significative. Au sein d'une même bande critique en revanche, le processus physiologique mis en jeu n'est que partiellement non additif et la fréquence de plus faible énergie aura tendance à être masquée.

L'interprétation proposée pour rendre compte de ces bandes de fréquence est qu'elles correspondent à une unité neurale sur la membrane basilaire (fig. 3.35). Ainsi, si deux fréquences sont dans une même bande, elles seront codées par les mêmes neurones et le processus sera donc additif.

Les travaux pionniers dans ce domaine ont été réalisés par Zwicker et ses collaborateurs (Zwicker *et al.*, 1957 ; Zwicker, 1961). Ces derniers définissent 24 bandes critiques. L'échelle issue de ces bandes critiques (numérotées de 1 à 24) est le Bark. La figure 3.35 présente le lien entre l'échelle en Barks et l'échelle fréquentielle. Au delà de 500 Hz, la relation est quasi-logarithmique. Néanmoins, la différence entre l'échelle logarithmique et l'échelle en Bark lors de l'utilisation d'un dispositif de substitution (The vOICe, Meijer, 1992) est nettement perceptible.

### Intérêt dans le cas du dispositif TheVIBE :

Nous avons vu que l'un des aspects essentiels pour l'établissement d'un couplage sensori-moteur à l'aide du dispositif TheVIBE est la capacité du système auditif à discriminer les différentes sources sonores qui composent le signal. Ces sources se différencient par leur fréquence et par leur azimut apparent.

En distribuant de manière équitable les niveaux des sources sonores dans les différentes bandes critiques de l'échelle en Bark, on minimise l'interaction entre ces différentes sources, non seulement en ce qui concerne la dimension fréquentielle (discrimination de deux sources par leur fréquence), mais aussi en ce qui concerne leur sonie (discrimination de la sonie de chaque source séparément).

Par ailleurs, dans la mesure où cette propriété est vraie pour chaque oreille individuellement, cette capacité à bien discriminer les écarts de sonie entre les différentes sources sonores améliore également la discrimination de l'IAD.

#### En pratique :

Dans la première expérience, l'échelle utilisée était tout simplement l'échelle de fréquence. Dans les expérimentations suivantes, nous avons utilisé une échelle en Barks.

Il n'y a pas de formule analytique exacte pour convertir l'échelle des fréquences en barks ( $z$ ). Une formule approximative est néanmoins fournie par Traunmüller (1990) :

$$z = \frac{26,81}{1 + 1960/f} - 0,53$$

Cette formule est inversible et permet de calculer une fréquence en fonction de l'échelle en Bark :

$$f = \frac{1960}{26.81/z - 1}$$

On déduit alors les fréquences permettant d'obtenir une répartition linéaire des niveaux en Bark. L'échelle utilisée s'étend de 300 Hz à 3000 Hz, soit de la bande  $z=3$  à la bande  $z=15$ .

### 3.4.7 Choix du nombre de champs récepteurs et du taux de rafraichissement de l'image

Ce n'est pas parce que l'on admet qu'un dispositif de substitution n'est pas un simple système de transfert de l'information mais un système de couplage sensori-moteur que la question du transfert de l'information doit être éludée.

Le nombre de champs récepteurs dans l'image et le taux de rafraichissement de l'activité de ces champs récepteurs détermine une certaine quantité d'information supposée être admise par le système sensitif de substitution. Néanmoins, la capacité en terme de transfert d'information d'un système sensitif n'est pas infinie : elle est bornée par ses caractéristiques notamment en terme de résolution spatiale (au sens de la dimension psychophysique concernée) et temporelle.

Dans cette section, nous allons donc nous pencher sur ces questions de capacité de transfert de l'information dans les systèmes sensoriels et sur les marges qu'elle impose quand au nombre de champs récepteurs et au taux de rafraichissement du signal sonore.

#### 3.4.7.1 La théorie de l'information

La méthode quantitative utilisée en psychophysique pour établir la quantité maximale d'information pouvant transiter dans un canal sensoriel est établie grâce aux outils de la "théorie de l'information".

Cette théorie, développée par Shannon et Weaver (1949) permet de quantifier la notion d'information. L'unité d'information dans cette théorie est le "bit". Il correspond à la quantité élémentaire d'information défini par la détermination d'une valeur parmi deux valeurs équiprobables : faire un choix entre 0 ou 1, pile ou face, c'est apporter au problème une quantité d'information égale à 1 bit, ou encore réduire de 1 bit l'incertitude liée à un problème.

La théorie de l'information utilise aussi la notion de "source". Une "source" d'information est une entité délivrant des "messages" définis par leur probabilité d'apparition (la nature sémantique du message n'est pas considérée). Par définition, la quantité d'information délivrée par un message dépend uniquement de sa probabilité d'apparition  $p$  selon la loi :

$$i(p) = -p \log_2(p)$$

Si  $p = 1/2$ ,  $i = 1$  bit.

La quantité d'information délivrée par la source est appelée entropie et correspond à la moyenne de la quantité d'information délivrée par ses messages. On a donc immédiatement :

$$H(s) = - \sum_i p_i \log_2(p_i)$$

La théorie de l'information utilise de plus la notion de "canal de transmission" au travers duquel les messages transitent avant d'être reçus par l'observateur. En général, ce canal n'est pas parfait : il va modifier les propriétés des messages et en brouiller la réception. Dans le cas d'une source à messages discrets, le canal sera défini par la donnée de la probabilité conditionnelle de recevoir le message  $R_j$  ( $1 < j < N_r$ ) parmi les  $N_r$  messages possibles, sachant que la source a envoyé le message  $S_i$  ( $1 < i < N_i$ ) parmi les  $N_i$  messages possibles.

$$\text{Canal} = \{ p(R_j/S_i), \text{ pour } 1 < i < N_i \text{ et } 1 < j < N_r \}$$

Ce brouillage conduit à une perte d'information. L'information transitant dans le canal malgré cette perte est l'information mutuelle ou "transinformation". On montre qu'elle correspond à



l'entropie de la source diminuée d'une quantité d'information liée au canal :

$$I(S,R) = H(R) - H(R/S)$$

La capacité du canal, enfin, correspond à la quantité maximale d'information susceptible d'y transiter. Cette capacité dépend uniquement des propriétés du canal. Pour l'atteindre en pratique toutefois, il peut s'avérer nécessaire de travailler sur les messages de la source. C'est ce que l'on appelle le "codage canal".

Nous allons voir dans quelle mesure ce formalisme appliqué à un système sensitif conçu comme un canal de transmission permet de dégager des bornes aux nombres de messages (c'est à dire au nombre de signaux élémentaires) admissibles, et de travailler sur les propriétés de ces messages (les caractéristiques des signaux élémentaires) pour améliorer la transmission.

### 3.4.7.2 Capacité de transfert globale des systèmes sensoriels

Une évaluation des capacités de transfert des systèmes sensoriels visuels, tactiles et auditifs est proposée dans Schmidt (1986, cité par Kaczmarek, 1995), sur la base des propriétés des neurones et de fibres afférentes pour chaque système. Elle aboutit respectivement à des valeurs de  $10^7$ ,  $10^6$  et  $10^5$  bits par secondes.

Si ce résultat permet de se faire une première idée des capacités relatives des systèmes sensoriels, il doit être considéré avec précaution. On pourrait par exemple penser que le système tactile est plus adapté à la substitution visuelle car sa capacité est meilleure que le système auditif. Cependant, il faut alors prendre en compte le fait que cette mesure évalue la capacité globale de tout le système sensitif : elle ne dit pas si et comment il est possible d'atteindre cette capacité. A titre d'illustration, la mobilisation de tout le système tactile nécessiterait l'utilisation d'une combinaison tactile intégrale, alors que le système auditif ne nécessite que de simples écouteurs.

Pour aller plus loin, nous devons donc nous pencher sur le transfert d'information dans les conditions particulières du dispositif de substitution que nous utilisons.

### 3.4.7.3 Nombre de champs récepteurs à partir des mesures psychophysiques de capacité de transfert par catégorisation

La mesure psychophysique de la capacité de transfert d'une dimension sensorielle utilise le modèle du canal discret décrit précédemment. La source d'information est ici le stimulus qui peut prendre un certain nombre de valeurs .

$$S = \{S_i, 1 < i < N_s\}$$

Ces valeurs sont réparties selon une échelle équi-JND que nous avons déjà mentionné, ce qui a pour but de maximiser le transfert d'information. Une explication qualitative est fournie dans Garner et Hake (1951). Je proposerai dans la suite du propos une approche quantitative de cette question.

Le récepteur du message est le sujet qui dispose également d'un nombre fini de réponses.

$$R = \{R_j, 1 < j < N_r\}$$

#### **Protocole :**

Le protocole expérimental consiste à demander au sujet d'attribuer à chaque stimulus la catégorie qui lui convient. Chaque stimulus est répété plusieurs fois de manière à disposer d'une information statistique. A l'issue de l'expérience, on dispose ainsi, pour chaque stimulus  $S_i$ , de la probabilité de



se voir attribuer la catégorie  $R_j$ , autrement dit, on dispose de l'ensemble des probabilités conditionnelles:

$$PC = \{ p(R_j/S_i), 1 \leq i \leq N_s, 1 \leq j \leq N_r \}$$

On dispose de plus de l'ensemble des probabilités concernant la réponse du sujet :

$$PR = \{ p(R_j), 1 \leq j \leq N_r \}$$

On est donc en mesure de calculer l'information mutuelle transitant par le canal sensoriel en question.

$$I(S,R) = H(R) - H(R/S)$$

soit en pratique :

$$I(S, R) = \sum_{i=1}^{N_s} \sum_{j=1}^{N_r} p(R_j/S_i) p(S_i) \log_2 \left( \frac{p(R_j/S_i)}{p(R_j)} \right)$$

L'expérience est répétée en augmentant peu à peu le nombre de valeurs du stimulus ( $N_s$ ) et le nombre de catégories ( $N_r$ ). Le lecteur intéressé par une explication plus développée de la méthode pourra se reporter à Bonnet (1986).

### Résultats :

Les courbes ci-dessous présentent les résultats obtenus avec des sons purs de différentes fréquences (fig. 3.35) ou de différents niveaux d'intensité sonore (fig. 3.36). On constate dans les deux cas une limite à la quantité d'information transmise. L'interprétation de cela est qu'à partir d'un certain nombre de stimulus envoyés au sujet, sa capacité à les catégoriser (à les reconnaître) sature.

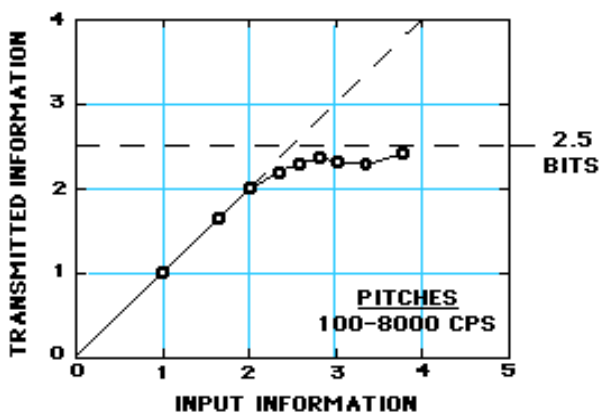


Figure 3.35 : mesure de la capacité de transfert du système auditif pour des sons purs de fréquence comprise entre 100 et 8000 Hz. Miller 1956 d'après Pollack, 1952.

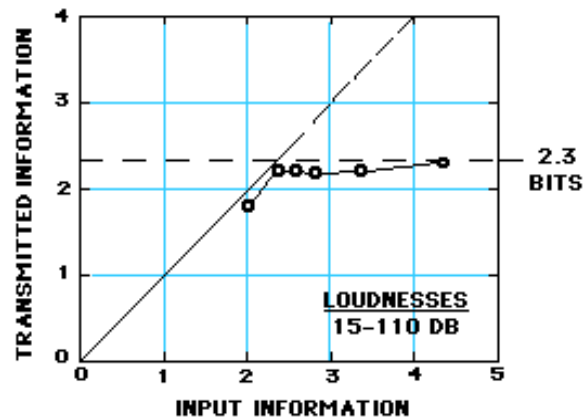


Figure 3.36 : mesure de la capacité de transfert du système auditif pour des sons purs d'intensité sonore comprise entre 15 et 110 dB. Miller, 1956, d'après Garner, 1953.

Concrètement, le sujet s'avère capable de reconnaître  $2^{2.3} = 5$  niveaux sonores et  $2^{2.5} = 7$  niveaux de fréquence. Si l'on augmente le nombre de stimulus d'entrée au delà de ces valeurs, cela n'augmente pas l'information transmise.

Application pratique :

Ces résultats nous donnent une première estimation du nombre de champs récepteurs admissibles. Selon cette approche, nous ne devons donc utiliser que 7 valeurs de fréquences : nous ne pouvons donc admettre que 7 niveaux de tonie différents

Par ailleurs, le codage du panoramique est réalisé à partir de l'IAD. Sachant que nous ne pouvons utiliser que 5 niveaux d'intensité sonore, nous ne pouvons admettre que 5 valeurs d'azimut : (-2,2), (-1,1), (0,0), (1,-1) ou enfin (2,-2).

En faisant l'hypothèse qu'il n'y a pas d'interaction entre sonie et tonie, le nombre de sources sonores est donc ici limité à

$$5 \times 7 = 35 \text{ sources sonores}$$

### **Mesure de la capacité sensorielle : le 7, un nombre magique?**

La valeur calculée pour le nombre de champs récepteurs paraît bien sur très faible et n'est pas en accord avec les valeurs choisies dans les expériences réalisées avec TheVIBE : dans Auvray, 2005, les auteurs utilisent 200 champs récepteurs. Nous allons donc nous pencher sur la signification de ces mesures et sur leur pertinence vis à vis de notre dispositif.

De nombreuses expériences de ce type ont été menées afin de déterminer la capacité d'un grand nombre de dimensions sensorielles : de la capacité de jugement du goût sucré ou salé à celle de discriminer la position d'un curseur sur une ligne horizontale, en passant par la capacité à discriminer des couleurs, des intensités sonores ou lumineuses, des fréquences sonores ou encore la taille de carrés sur un écran. Ces mesures sont surprenantes à bien des égards, et ont fait couler beaucoup d'encre.

Etant donné la variété des dimensions considérées, tout d'abord, on aurait pu s'attendre à ce que leurs capacités de transferts soit très différentes. Dans la pratique, ce n'est pas le cas. Dans un article intitulé "The magical number seven, plus or minus two : some limits on our capacity for processing information", Miller (Miller, 1956) montre que, toutes dimensions confondues, la distribution des valeurs de capacité de transfert présente une moyenne de 2,6 bits et un écart -type de 0,6 bits, soit  $7 \pm 2$  catégories. Cette valeur est valable aussi bien pour les continuums prothétiques que pour les continuums métathétiques (voir par exemple les figures 3.35 et 3.36), ce qui est particulièrement contre-intuitif.

Cette capacité moyenne, ensuite, paraît singulièrement faible, et ne semble pas rendre compte de la finesse du vécu sensitif concret. Si, par exemple, l'oreille humaine ne peut construire que 7 catégories sur la base de la fréquence d'un son, on est en droit, par exemple, de s'interroger sur la capacité d'un pianiste à distinguer les 88 notes de son instrument.

Enfin, cette valeur est également, de manière surprenante, robuste à la marge, c'est à dire à l'étendue des valeurs possibles du stimulus. Une expérience de Luce *et al.* (1976) sur la sonie d'un son pur montre que le nombre de catégories plafonne à 7 et ce que l'on travaille sur une plage de 20dB ou de 40 dB.

Une hypothèse concernant cette limite est bien sûr celle de l'empan mnésique, dont la valeur est du même ordre de grandeur. Miller toutefois montre dans son article que cette hypothèse est loin d'être démontrée et préfère conclure à une coïncidence.

Pour conclure, revenons sur la tâche effectuée par le sujet. Celui-ci doit attribuer à chaque stimulation un nombre et ce de manière reproductible, c'est à dire qu'il doit reconnaître chaque stimulation. Si cette condition est suffisante pour que le sujet soit capable de différencier deux

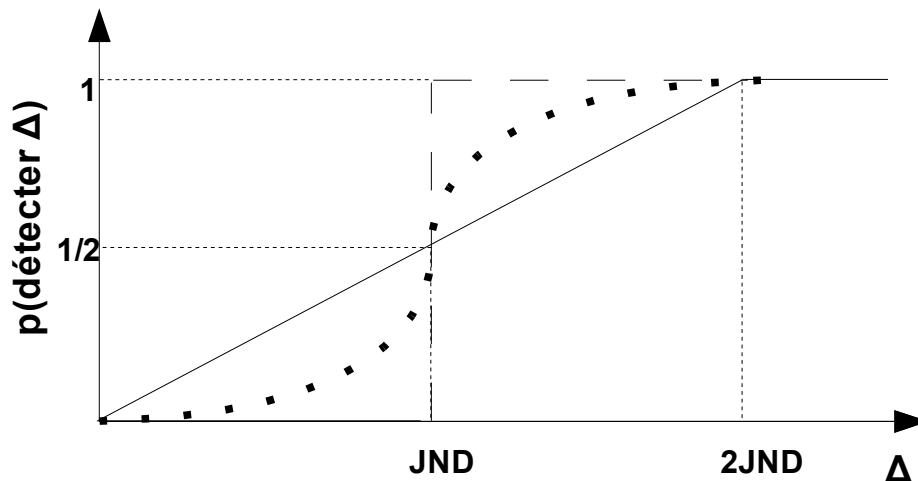


Figure 3.37 : Forme typique d'une courbe de mesure du JND (en points). La courbe en trait plein illustre la courbe obtenue par un modèle de canal à bruit aléatoire uniforme proposé dans cette section. Celle en points illustre le modèle considéré par Zwicker (1975)

stimulations, elle n'est pas nécessaire à cela. L'établissement d'un couplage sensori-moteur nécessite que le sujet soit capable de percevoir les changements entre deux stimulations, mais pas nécessairement qu'il soit capable de reconnaître chaque stimulation prise individuellement. Le nombre de champs récepteurs calculé ici est donc vraisemblablement une borne inférieure. Nous allons maintenant nous pencher sur une méthode permettant de quantifier la capacité d'un canal en se basant uniquement sur les capacités du sujet à discriminer une variation du signal.

#### 3.4.7.4 Nombre de champs récepteurs à partir des mesures psychophysiques de la différence juste notable

La mesure psychophysique liée à la capacité d'un sujet de discriminer deux niveaux de stimulation est la différence juste notable (JND). Cette grandeur caractéristique correspond à l'écart entre deux niveaux du stimulus tel que la probabilité de détecter une différence entre les deux stimulus est de 1/2. En pratique, on présente au sujet des signaux deux à deux avec des écarts différents et on mesure la probabilité qu'a le sujet de détecter une différence en fonction de l'écart. Le type de courbe obtenu est présenté figure 3.37.

Intuitivement, cette mesure nous donne une nouvelle estimation du nombre maximal de niveaux admissible sur une marge sensorielle donnée : dans le cas où le JND est constant sur l'échelle psychophysique considérée (voir § 3.3.1), ce nombre est alors de l'ordre de la marge divisée par le JND (Zwicker, 1975, cité dans Capelle *et al.*, 1998). De la sorte, chaque stimulus est à peu près différentiable de son voisin.

L'objectif de ce paragraphe est d'étudier cette question quantitativement à l'aide des outils de la théorie de l'information. Il s'agit en particulier de voir dans quelle mesure l'information que l'on ajoute en ajoutant un niveau de stimulation de plus ne compense pas la perte d'information liée à l'imprécision que l'on induit alors sur l'ensemble des niveaux.

Dans un premier temps, nous nous pencherons sur la question de l'échelle équi-JND proposée par Garner et Hake (1952). Nous verrons ensuite ce qui se produit lorsque l'on ajoute des niveaux sur ce type d'échelle.

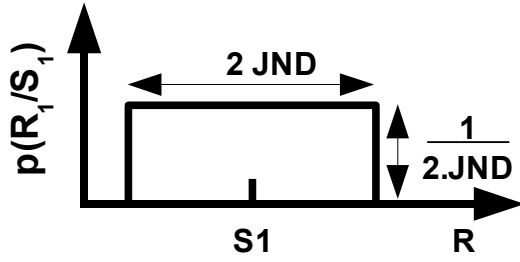


Figure 3.38 : Modèle du canal de transmission du système sensitif considéré

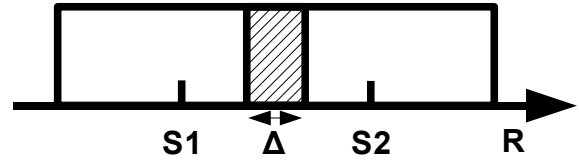


Figure 3.39 : Cas où deux signaux sont séparés par une distance inférieure au JND. La zone hachurée correspond au cas où S1 et S2 ne sont pas discriminés.

### Modélisation du problème :

On modélise le système sensitif par un canal à bruit uniforme additif. Soit  $S_i$  le signal envoyé, le signal "reçu" par le système sensitif s'écrit :

$$R_i = S_i + b$$

où  $b$  est une variable aléatoire de distribution uniforme comprise entre  $-JND$  et  $+JND$  (fig 3.38). Dans le cas où deux signaux  $S_1$  et  $S_2$  sont séparés d'une distance inférieure à  $2 JND$ , les réponses  $R_1$  et  $R_2$  se superposent (fig 3.39). Dans la zone où les réponses se superposent, la probabilité de répondre  $R_1$  est égale à la probabilité de répondre  $R_2$  et vaut  $1/2$ . La probabilité conditionnelle de  $R_1$  sachant  $S_1$  vaut donc maintenant :

$$p(R_1/S_1) = 1 - \frac{1}{2} \times \frac{\Delta}{2.JND}$$

, avec  $\Delta = 2.JND - (S_2 - S_1)$  d'où

$$p(R_1/S_1) = 1 - \frac{1}{2} \times \left(1 - \frac{S_2 - S_1}{2.JND}\right)$$

Lorsque  $S_1 = S_2$ , les probabilités conditionnelles de  $R_1$  et  $R_2$  valent  $1/2$ , c'est à dire que les signaux sont parfaitement non discriminable. Lorsque  $S_1$  et  $S_2$  sont séparés de  $2 JND$ , les probabilités conditionnelles de  $R_1$  et  $R_2$  valent toutes les deux 1 : les signaux sont parfaitement discriminable. Entre les deux, l'évolution est linéaire (fig. 3.37).

### Répartition des niveaux sur une échelle sensorielle :

Pour étudier la question de la répartition des niveaux de stimulation sur le continuum sensoriel, nous allons nous placer dans le cas proposé par Garner et Hake : les niveaux sont répartis à intervalles réguliers sur une échelle psychophysique équi-JND.

On considère pour cela trois niveaux séparés chacun de  $2.JND$  : les sources sont toutes parfaitement différenciables (fig. 3.40, cas 1). Le canal est donc défini par  $p(R_j/S_i) = \delta(i,j)$ , soit sous forme matricielle

$$C_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

On considère par ailleurs que les messages  $S_i$  sont équiprobables :  $p(S_i) = 1/3$  quelque soit  $i$ .

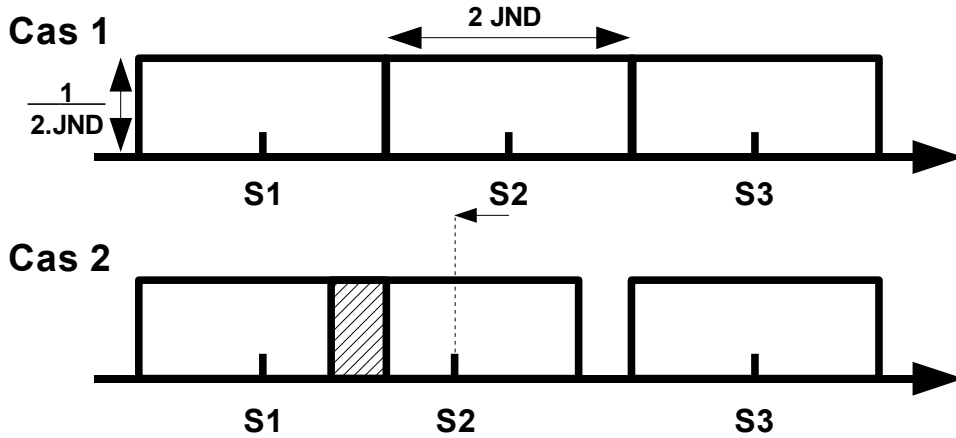


Figure 3.40 : Etude de l'échelle équi-JND proposée par Garner et Hake (1952)

Il en est de même pour la probabilité de réception des messages  $p(R_i)$ . En effet, d'après le théorème des probabilités totales :

$$p(R_i) = \sum_{j=1}^3 p(R_i/S_j) p(S_j) = \frac{1}{3}$$

En utilisant la formule de la transinformation proposée plus haut, on obtient :

$$I(S, R) = \sum_{i=1}^{N_s} \sum_{j=1}^{N_r} \frac{1}{3} \delta(i, j) \log_2 \left( \frac{\delta(i, j)}{1/3} \right) = \log_2(3)$$

Dans ce cas, la transinformation est égale à l'entropie de la source : toute l'information est transmise.

Voyons maintenant ce qui se produit si l'on déplace le niveau de la source  $S_2$ , par exemple vers  $S_1$  (fig. 3.40, cas 2). Le canal de transmission s'écrit alors :

$$C_2 = \begin{bmatrix} 1 - \delta & \delta & 0 \\ \delta & 1 - \delta & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ avec } \delta = \frac{1}{2} \left( \frac{\Delta}{2 \cdot \text{JND}} \right)$$

La figure 3.41 représente l'évolution de la transinformation dans le canal  $C_2$  en fonction du décalage  $\Delta$ . Pour faciliter la lecture du graphique, la courbe représente la fonction  $2^{I(S,R)}$  qui correspond à la transinformation exprimée en nombre de niveaux transmis.

On voit que pour  $\Delta=0$ , les trois niveaux sont transmis par le canal. La quantité d'information diminue ensuite jusqu'à un minimum de 1.88 catégories lorsque les niveaux  $S_1$  et  $S_2$  sont confondus.

Le résultat est généralisable de proche en proche et permet de conclure que l'échelle permettant de transmettre le mieux l'information est une échelle où les niveaux sont répartis en fonction de la valeur du JND, ou de manière équivalente, s'il sont répartis uniformément sur une échelle équi-JND. L'hypothèse de Garner et Hake est donc vérifiée dans notre modèle.

#### Nombre de niveaux admissibles sur une échelle équi-JND :

Le cas étudié précédemment est relativement immédiat de sorte qu'un développement mathématique peut paraître inutile. Il va néanmoins nous permettre d'aborder maintenant une question beaucoup moins immédiate : combien de niveaux doit-on mettre sur une échelle équi-

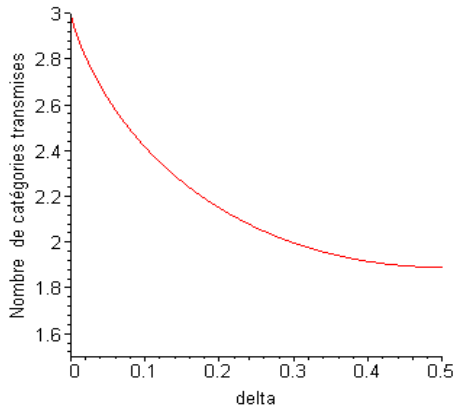


Figure 3.41 : évolution de la transinformation lorsque l'on déplace la position d'une source dans une échelle équi-JND.

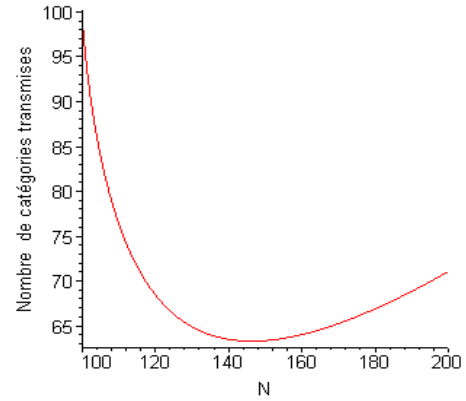


Figure 3.42 : évolution de la transinformation dans un canal équi-JND en fonction du nombre de niveaux.

JND ? Cette fois en effet, il y a compétition entre deux processus. D'un côté, chaque niveau ajouté réduit l'écart entre tous les niveaux et augmente l'incertitude sur leur détermination, donc diminue la transinformation. De l'autre, chaque niveau ajouté augmente la quantité d'information envoyée par la source. Lequel des deux processus est-il prépondérant ?

On considère donc  $N$  niveaux répartis à intervalles réguliers sur une marge  $M$  (fig. 3.43).

Bien sûr, selon notre modèle, tant que le nombre de niveaux est inférieur à  $M/(2 \cdot \text{JND})$ , les sources sont entièrement discriminables, le canal est parfait et la quantité d'information transmise est  $\log_2(N)$ . On pose donc  $N_0 = M/(2 \cdot \text{JND})$ . Calculons maintenant la transinformation pour  $N > N_0$ .

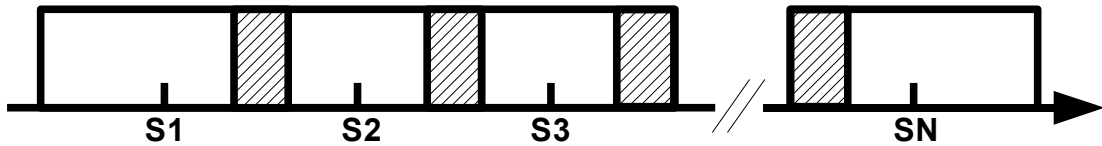


Figure 3.43 : Cas d'un canal équi-JND avec  $N > M/(2 \cdot \text{JND})$

La matrice du canal s'écrit maintenant :

$$CN = \begin{pmatrix} 1 - \delta & \delta & 0 & 0 & 0 & \dots & 0 \\ \delta & 1 - 2\delta & \delta & 0 & 0 & \dots & 0 \\ 0 & \delta & 1 - 2\delta & \delta & 0 & \dots & 0 \\ 0 & 0 & \delta & 1 - 2\delta & \delta & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \delta & 1 - 2\delta & \delta \\ 0 & 0 & 0 & \dots & 0 & \delta & 1 - \delta \end{pmatrix}$$

La transinformation s'écrit alors :

$$I(N) = \log_2(N) + \frac{1}{N} [2(N-1)\delta \log_2(\delta) + 2(1-\delta)\log_2(1-\delta) + (N-2)((1-2\delta)\log_2(1-2\delta))] ]$$

Par ailleurs, on a :

$$\Delta = 2 \text{ JND} - \frac{M}{N} = 2 \text{ JND} \left(1 - \frac{N0}{N}\right)$$

d'où

$$\delta = \frac{1}{2} \times \frac{\Delta}{2 \cdot \text{JND}} = \frac{1}{2} \left(1 - \frac{N0}{N}\right)$$

L'évolution de l'information mutuelle en fonction du nombre de niveaux choisis est présentée fig. 3.42. La courbe est calculée pour  $N0 = 100$ . Le modèle est valide entre 100 et 200 catégories seulement. Au delà, il y a interaction entre les sources  $n$  et  $n+2$  et notre modèle n'est donc plus valide. On constate que la perte d'information n'est pas compensée par le gain obtenu lorsque l'on rajoute des niveaux. Selon ce modèle et dans le cas présenté, le nombre optimal de niveaux est donc  $N0$ . Le même type de courbe est obtenu pour des valeurs de  $N0$  allant de 3 à 2000.

On remarquera toutefois que le choix de  $N0 = M/(2 \cdot \text{JND})$  est sans doute sous évalué, car en pratique, les capacités de discrimination du système sensitif sont meilleures que celle calculées par notre modèle. En effet, les courbes obtenues lors des mesures de JND ne sont pas linéaires mais sigmoïdales (fig. 3.37). Quoiqu'il en soit, le nombre de niveaux doit être compris entre  $N0$  et  $2N0$ , cette dernière valeur correspondant au cas où l'on considère un modèle en escalier ou la probabilité de détecter un changement vaut 0 en dessous du JND et 1 au dessus.

#### Application numérique :

Sonie : le JND pour une échelle de sonie en dB est à peu près constant et de l'ordre de 0.6 dB (fig. 3.28). Si l'on travaille sur une plage de 60 dB, on a donc  $N0 = 60/1.2 = 50$  niveaux, on peut donc admettre  $N_s$  niveaux de sonie tel que :

$$50 < N_s < 100$$

Rq : cette valeur est à comparer à la précision obtenue sur l'activité du champ récepteur en fonction du nombre d'échantillons choisis. Pour 10 échantillons, cette précision est de l'ordre de 8 niveaux sur 255, ce qui donne  $255/8 = 31$  niveaux. A priori, une partie de l'incertitude associée au sous-échantillonnage est donc détectable. Ce nombre est donc un peu faible de ce point de vue.

Pour l'azimut : le JND pour une échelle de sonie en dB est à peu près constant et de l'ordre de 0.6 dB (fig 3.28). Si l'on travaille sur une plage de 12 dB pour le codage de l'IAD, on a  $N0 = 12/1.2 = 10$  niveaux, on peut donc admettre  $N_p$  niveaux tels que :

$$10 < N_p < 20$$

Pour la fréquence : sur la bande 300-3000 Hz, le JND relatif  $\Delta(x)/x$  est à peu près constant et égale à 0.2% (fig. 33). Calculons le JND absolu sur une échelle logarithmique.

Par définition, sur une échelle logarithmique,

$$\text{JND} = \log(x + \Delta(x)) - \log(x)$$

d'où

$$\text{JND} = \log(1 + \Delta(x)/x) = 0.002$$

Sur la bande 300-3000hz, on a donc :

$$N0 = (\log(3000) - \log(300)) / (0.004) \sim 570 \text{ niveaux}$$

et donc un nombre  $N_f$  de fréquences admissibles :

$$570 < N_f < 1140$$

Le nombre total de sources sonores admissible serait donc compris entre 5000 et 20000 sources.

Le nombre obtenu ici est cette fois très élevé par rapport aux valeurs utilisées dans la pratique. La raison est que le nombre de fréquences admissibles ne tient pas compte du phénomène de masquage. En effet, si deux sources sonores de fréquences voisines sont discriminables lorsqu'elles sont présentées l'une après l'autre, elles ne le sont pas forcément lorsqu'elles sont présentées ensemble.

### 3.4.7.5 Nombre de champs récepteurs en prenant en compte l'existence de bandes critiques

Le phénomène de masquage a, comme nous l'avons vu, donné naissance à la notion de bandes critiques. Deux fréquences appartenant à une même bande auront tendance à interagir. Le sujet ne discerne plus alors qu'une seule fréquence dans le signal.

En revanche, si les sources sont dans des bandes critiques différentes, elles seront discriminables dans le signal. De plus, leurs intensités sonores ne s'additionneront pas, et les valeurs d'ILD en particulier ne se mêleront pas. En conséquence, le nombre de fréquences admissibles est de l'ordre du nombre de bandes critiques dans la bande de fréquence considérée.

Application numérique :

En utilisant la formule permettant de calculer le numéro de la bande critique en Bark présentée au paragraphe 3.3.3, on obtient :

$$z(300) \sim 3 \text{ et } z(3000) \sim 16$$

Entre 300 et 3000 Hz, il y a donc 13 bandes critiques.

En considérant qu'au sein d'une même bande critique, les sources sonores de différentes valeurs d'IAD n'interagissent pas et sont donc discriminables simultanément (la validité de cette hypothèse a été étudiée au paragraphe 3.3.2), et en utilisant le résultat précédent à ce sujet ( $10 < N_p < 20$ ), on obtient donc

$$130 < N < 260$$

Dans la pratique, c'est la valeur  $N = 200$  que nous avons utilisé dans toutes les expérimentations

### 3.4.8 Aspects temporels : choix de la fréquence de rafraichissement de la caméra et de la durée des échantillons sonores

Nous allons pour conclure nous pencher sur trois paramètres définis dans le fichier d'option du logiciel :

- 1- La taille du buffer mémoire sonore, c'est à dire le nombre d'échantillons utilisés pour



construire une "image sonore", et la fréquence d'échantillonnage sonore. Ces deux paramètres conditionnent directement la durée d'une "image sonore". La question qui se pose est alors de savoir s'il y a une durée minimale de stimulation pour évoquer une sensation auditive.

2- Le taux de rafraichissement des images de la caméra. La question qui se pose est de savoir jusqu'à quelle échelle le système auditif est encore capable de détecter des changements temporels.

### **Durée du buffer sonore :**

Fish (1976) et Meijer (1992) posent tous deux la question de la durée optimale pour une stimulation élémentaire. Dans le cas du dispositif de Fish, il s'agit de faire en sorte que le signal soit composé d'un nombre suffisant de cycles pour pouvoir être perçu. En rapport avec la vitesse du spot lumineux qui balaye l'image de haut en bas et de droite à gauche, et qui conditionne donc la durée du signal associé à chaque position du spot, il en déduit une fréquence minimale pour le codage sonore.

Dans le cas du dispositif de Meijer, la durée du buffer sonore correspond au temps mis pour balayer l'image de gauche à droite divisé par le nombre de colonnes dans l'image (codées par des sons séparés). Meijer s'interroge alors sur les composantes fréquentielles induites par le fait que le son est présenté dans une fenêtre temporelle de courte durée. En effet, ce fenêtrage introduit des artefacts fréquentiels qui dépendent de la taille de la fenêtre. Il en déduit l'écart minimal  $\Delta f$  entre deux niveaux d'une même colonne pour que deux fréquences associées à deux rangées voisines ne puissent pas être confondues avec les fréquences liées au fenêtrage temporel.

Dans le cas des dispositifs où l'image est codée en simultané (sans balayage temporel), comme TheVIBE ou le PSVA, s'interroger sur la durée minimale de stimulation en rapport avec la perception auditive, c'est en fait omettre la composante motrice du dispositif vu comme système de couplage sensori-moteur. En effet, tout comme dans le cas du système visuel, ce sont les mouvements de la caméra qui vont déterminer la durée de fonctionnement des différentes sources sonores. Le sujet peut donc adapter les mouvements de sa tête de manière à pouvoir sentir distinctement les sons en provenance du dispositif.

Pour cela il s'agit néanmoins de s'assurer de la continuité du signal d'une image à une autre, en particulier de faire en sorte que la phase de tous les signaux élémentaires à la fin du buffer "n" correspondent à la phase de tous les signaux élémentaires au début du buffer "n+1". Pour cela, le calcul du son dans TheVIBE utilise un marqueur temporel absolu : le calcul du signal pour l'image "n" commence au temps  $t = t_n$  avec

$$t_n = n \cdot \Delta t$$

avec  $\Delta t = \frac{N_s}{f_e}$  où  $N_s$  et  $f_e$  sont respectivement la taille du buffer mémoire (en nombre d'échantillons) et la fréquence d'échantillonnage du signal sonore.  $\Delta t$  correspond alors à la durée du buffer sonore.

En pratique, nous avons déterminé la taille du buffer en fonction du taux de rafraichissement de l'image, de sorte que la durée de présentation d'une image sonore corresponde à la durée d'affichage d'une image, c'est à dire à la période de rafraichissement de la caméra.

### **Taux de rafraichissement de la caméra :**

S'il n'y a pas de limite *a priori* à la durée de présentation d'un image pour qu'elle soit interprétée, il n'y a *a priori* aucune limite sur la fréquence de rafraichissement de la caméra vidéo. En revanche,

il est vraisemblable qu'à partir d'une certaine fréquence, le rafraîchissement de la caméra soit inutile au sens où il ne puisse pas être utilisé par le système auditif. Dans le cas du système visuel, on considère généralement qu'une cadence de 25 images par secondes suffit à donner à un film une fluidité acceptable pour l'œil, et que des cadences plus élevées n'améliorent pas significativement la qualité de la stimulation. On remarquera toutefois que cette valeur concerne une stimulation destinée à "être vue" et non une stimulation destinée à voir, comme c'est le cas d'une prothèse visuelle. Néanmoins, cette valeur nous donne un ordre de grandeur de la résolution temporelle du système visuel.

De la même manière, on peut donc s'interroger sur les ordres de grandeur de la résolution temporelle de l'oreille. Tout comme il existe une fréquence de fusion au-delà de laquelle le scintillement d'un stimulus visuel n'est plus perçue, il existe une limite, appelée Auditory Fusion Threshold (AFT), au-delà de laquelle deux stimulus auditifs présentés successivement semblent confondus (Miller, 1948). Un tableau des valeurs standard d'AFT pour différents âges est fourni figure 3.44.

**Table 1**

<b>MEAN AND STANDARD DEVIATION OF NORMAL AFT<sup>THRESHOLD</sup> AND SUGGESTED CUT-OFF SCORES</b>			
<b>AGE (Years)</b>	<b>Normal Mean (msec)</b>	<b>Std. Dev.</b>	<b>Cut off*</b>
3	23	17	57
4	17	6	29
5	13	5	23
6	12	4	20
7	9	4	17
8-50	8	3	14
60	9	8	25
70	22	17	56

Figure 3.44 : Tableau des normes standards de seuil de détection temporel auditif en fonction de l'âge (Keith, 2001, d'après Mac Croskey et Keith, 1996)

L'AFT moyen pour un adulte est d'environ 8 msec, soit une fréquence de fusion équivalente de 125 Hz. Le système auditif semble donc posséder une bien meilleure résolution temporelle que le système visuel, dont la fréquence de fusion est de l'ordre de la dizaine de Hz (Hecht et Verrijs, 1933).

Une information plus fine au sujet de la résolution temporelle du système auditif est disponible via la mesure de "fonctions de transfert" psychophysiques. Le principe de la mesure consiste à présenter au sujet un signal (en général un bruit blanc), modulé en amplitude par une sinusoïde. On mesure alors la capacité du sujet à déterminer si le signal est modulé ou non, en fonction de la fréquence et du coefficient de modulation (le rapport entre l'amplitude de la modulation et l'amplitude du signal). Les résultats sont présentés sous la forme d'une courbe donnant, en fonction de la fréquence, le coefficient de modulation à partir duquel le sujet distingue un son modulé d'un son non-modulé.

La figure 3.45 présente les mesures effectuées par Viemeister (Viemeister, 1979) en utilisant des bruits blancs dans trois bandes de fréquences : 200, 2000 et 10000 Hz. On constate que les courbes sont relativement similaires quelque soit la fréquence. La résolution temporelle du système auditif est constante jusqu'à environ 8 Hz puis décroît progressivement jusqu'à environ 1000 Hz.

Pour le cas de TheVIBE, la courbe la plus représentative est celle calculée à partir d'un bruit

blanc dans une bande autour de 1000 Hz. Dans ce cas, la fréquence de coupure est d'environ 50 Hz.

En conclusion, on peut donc dire que la fréquence de 25 images par secondes est acceptable, au moins en lien avec les ordres de grandeur concernant le système visuel. Augmenter cette fréquence pourrait toutefois augmenter l'efficacité du dispositif, vraisemblablement jusqu'à des valeurs de l'ordre de 50 Hz. Au delà, le transfert d'information sera limité par la résolution temporel du système auditif.

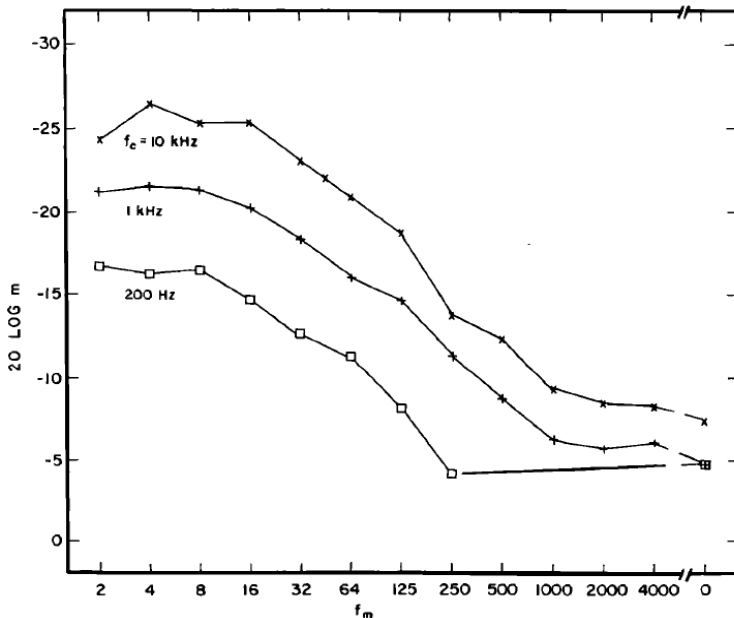


Figure 3.45 : fonction de transfert temporelle du système auditif pour des bruits blancs de différentes fréquences modulés en amplitude (Viemeister, 1979).

En pratique :

La fréquence d'échantillonnage du signal sonore est déterminée par la bande de fréquence que l'on souhaite utiliser. Toujours d'après le théorème de Shannon (voir §2.1), un signal sonore échantillonné à une fréquence  $f_e$  pourra transmettre la bande de fréquence  $[0, f_e/2]$ .

Dans la mesure où l'on travaille dans la bande 300-3000 Hz, un taux d'échantillonnage de 22100 Hz est largement suffisant. On a donc :

$$f_e = 22100 \text{ Hz}$$

Par ailleurs, on souhaite que la durée du buffer sonore soit de l'ordre de la période de rafraichissement de la caméra. Cette période est égale à

$$T_e = 1/25 = 40 \text{ ms}$$

Le nombre d'échantillons pour obtenir un son de 40 ms avec une fréquence d'échantillonnage de 22100 Hz est de

$$N_e = f_e * T_e = 884 \text{ échantillons}$$

En pratique, nous avons utilisé un buffer de 1024 échantillons, ce qui assure une continuité entre les différents signaux sonores (pas d'attente du signal de la caméra) et évite d'éventuels silences.

### 3.4.9 Choix de la fonction de conversion visuo-sonore

La dernière étape de la configuration du logiciel consiste à déterminer le lien entre l'espace de l'image, représenté par les  $N$  valeurs d'activité des champs récepteurs, et l'espace sonore, formé par les  $N$  sources sonores. Comment doit on mettre concrètement en correspondance les coordonnées spatiales des champs récepteurs et les caractéristiques psychophysiques (azimut, tonie) des sources sonores qui leur sont respectivement associées?

#### 3.4.9.1 Critères généraux

Dans leur article, Capelle *et al.* (1998) tentent d'identifier des règles générales concernant cette mise en correspondance. Selon lui,

1. Le déplacement d'un objet dans une scène visuelle devrait conduire à des signaux sonores voisins<sup>6</sup>
2. Les différents niveaux de résolution issues de leur modèle de vision [basée sur plusieurs niveaux d'échantillonnages spatialement variants] doit correspondre à des groupes spécifiques de fréquences de manière à faciliter l'évaluation des dimensions de l'objet.<sup>7</sup>

La deuxième remarque ne s'applique en pratique qu'aux dispositifs munis de différents niveaux de résolution. Elle me paraît de ce fait difficilement généralisable.

En revanche, la première remarque introduit de manière très intéressante la notion de voisinage. Cette hypothèse que deux images "voisines" doivent donner lieu à deux "sons voisins" repose toutefois sur un certain nombre de pré-requis que j'ai tenté d'identifier. J'ai ensuite tenté d'en déduire des contraintes sur le mapping.

#### Modélisation du problème :

La notion de voisinage pour des images ou pour des sons est bien sûr très général. Nous devons donc en proposer une définition. Pour cela nous allons nous placer dans le cas unidimensionnel. On considère d'une part une image constituée d'une ligne composée de  $N$  pixels  $p_i$   $i \in [1..N]$ , ordonnés par exemple de la gauche à la droite. On considère d'autre part  $N$  sources sonores de tonie (ou tout autre paramètre psychophysique)  $t_k$ ,  $k \in E [1..N]$  toutes discriminables entre elles, ordonnées également selon la dimension psychophysique considérée.

La question qui se pose est de savoir comment on doit associer les  $N$  pixels  $p_i$  aux  $N$  sources sonores  $t_k$ .

Dans le cadre de la théorie sensori-motrice de la perception, nous proposons la définition suivante : deux images sont voisines si et seulement-ci elles correspondent à deux positions voisines de la caméra.

On considère alors le cas particulier d'un univers visuel composé d'une seule source lumineuse ponctuelle, c'est à dire qui n'active qu'un seul pixel de l'image à la fois, en l'occurrence le pixel  $j$ . Le fait d'utiliser un cas particulier ne remet en rien en cause la portée générale du propos car ce qui doit être vérifié dans tous les cas doit l'être notamment dans ce cas.

Dans le cas considéré, la valeur des pixels de l'image s'écrit  $p_i = \delta(i, j)$ . On appellera  $I$  l'ensemble de toutes les images ainsi obtenues et  $im(j)$  l'élément de cet espace correspondant à l'image :

<sup>6</sup> "displacements of an object in a visual scene should lead to neighbor auditory signals", p1283

<sup>7</sup> "Differents levels of resolution in the model of vision should correspond to specific sets of tones, in order to facilitate evaluation of object dimension and localization", p1283

$$im(j) = im \in I \text{ tq } \{p_i = \delta(i,j), i \in [1..N]\}$$

On considère un déplacement élémentaire de la caméra comme le passage du point du pixel  $j$  au pixel  $j+1$  (ou de manière réversible du pixel  $j+1$  au pixel  $j$ ). Un déplacement plus petit n'aurait pas de sens du point de vue sensori-moteur puisqu'il ne provoquerait pas de variation sensitive. On définit donc un voisinage dans l'espace des images  $I$  de la manière suivante :

$$im(i) \text{ et } im(j) \text{ sont voisines si et seulement si } |i-j| < 2$$

De la même manière, on peut définir un voisinage dans l'espace sensoriel comme la plus petite variation entre deux signaux de l'ensemble des signaux utilisés entraînant un changement perceptible dans les propriétés du son produit.

On a alors :

$$t_k \text{ et } t_l \text{ sont voisins si et seulement si } |l-k| < 2$$

On souhaite que deux images voisines correspondent à deux sons voisins. Soit  $im(i)$ ,  $im(j)$  deux images,  $t_k$  et  $t_l$  deux sources sonores, cette contrainte peut s'écrire

$$|i-j| < 2 \Rightarrow |l-k| < 2 \quad (1)$$

On considère maintenant l'application  $f$  qui, à l'index  $i$  de chaque pixel associe l'index  $k$  de la source associée

$$f : [1..N] \rightarrow [1..N]$$

$$i \rightarrow k$$

Montrons que si  $f$  est bijective, la condition (1) est vraie si et seulement si  $f$  est strictement monotone.

Pour cela montrons par récurrence que si  $f$  est bijective et que (1) est vraie, alors  $f$  est strictement monotone

1- On pose  $f(1) = m$

2- D'après (1),  $f(2)$  peut alors valoir  $m$ ,  $m-1$  ou  $m+1$ . Sachant de plus que  $f$  est bijective,  $f(2) = m$  est impossible.

3- Si  $f(2) = m+1$ , par le même raisonnement on déduit  $f(3) = m+2$  ou  $f(3) = m$ . Sachant que  $f$  est bijective, la valeur  $f(3) = m$  est impossible. On en déduit  $f(3) = m+2$ . Un raisonnement similaire dans le cas où  $f(2) = m-1$  aboutit à  $f(3) = m-2$

Ainsi si  $f$  est bijective et vérifie (1), alors  $f$  est soit strictement croissante soit strictement décroissante

Montrons maintenant que si  $f$  est bijective et strictement monotone alors (1) est vraie. Pour plus de simplicité, on raisonnera sur le cas strictement croissant. Le cas strictement décroissant est obtenu par un raisonnement similaire.

1- Si  $f$  est strictement croissante, on a  $f(i+1) - f(i) \geq 1$  et donc par extension  $f(N) - f(1) \geq N-1$ . Dans la mesure où  $\max(f) = N$  et  $\min(f) = 1$ , on en déduit  $f(N) = N$  et  $f(1) = 1$ .

Sachant qu'il reste  $N-2$  sources sonores à attribuer aux  $N-2$  pixels de l'image, on en déduit  $f(i+1) - f(i) = 1$  quel que soit  $i$ . Par récurrence, on obtient alors immédiatement

$$f(i) = i$$

Si  $f(i) = i$  alors la propriété (1) est vraie. CQFD.

### Conclusion :

A l'issue du travail ci-dessus, on peut identifier un certain nombre de possibilités concernant le mapping visuo-sonore. Si on considère une image formée de  $N$  pixels, plusieurs possibilités se présentent, en fonction du nombre de sources sonores différentes dont on dispose et des propriétés que l'on souhaite obtenir :

1. On ne dispose que d'une seule source sonore. Dans ce cas, l'application  $f$  est constante : tous les pixels vont faire varier l'intensité de la même source sonore.
2. On considère  $L$  sources sonores (différentes) avec  $L < N$ . Dans ce cas, l'application  $f$  n'est pas injective, mais il est difficile d'en dire plus.
3. On considère  $N$  sources sonores toutes différentes. On peut alors définir une fonction de mapping bijective entre l'espace des pixels et l'espace des sources sonores. Sans autre contraintes, un tirage aléatoire sans remise est admissible.
4. On considère  $N$  sources sonores toutes différentes et on considère en plus la propriété concernant le voisinage défini plus haut. Dans ce cas, la fonction de mapping bijective doit être strictement monotone selon chaque dimension de l'image.

Il est intéressant de noter qu'aucune de ces conditions ne supposent l'existence d'une "métrique sensorielle". Le cas 3 nécessite l'utilisation d'une échelle de catégorie, c'est à dire que chaque source sonore est différente de l'autre et donc possède un "nom" différent. Le cas 4 impose d'ajouter la notion d'ordre pour définir la monotonie et la notion de voisinage, mais pas la notion de norme, une échelle ordinale des valeurs suffit donc.

Le fait de définir la notion de voisinage par le mouvement de la caméra possède une autre propriété. Dans le cas unidimensionnel, ce mouvement est entièrement contraint (c'est un mouvement de translation selon la dimension considérée). Si maintenant on travaille sur une image bidimensionnelle, plusieurs possibilités sont envisageable.

En particulier, on peut considérer que les deux dimensions de l'image sont les axes cartésiens  $x$  et  $y$ . Dans ce cas, un mouvement élémentaire correspondra à une translation de la caméra sur chaque axe. La fonction de mapping bidimensionnelle sera donc strictement monotone selon ces deux axes. C'est ce qu'on appellera le mapping euclidien.

On peut néanmoins considérer un autre répertoire de mouvements élémentaires de la caméra : celui des rotations et des zooms. La fonction de mapping devra alors être strictement monotone si on repère la position du champ récepteur par ses coordonnées polaires  $(\rho, \theta)$ . C'est ce qu'on appellera un mapping polaire.

On remarquera enfin que cette propriété de monotonie est une propriété nécessaire, mais pas forcément suffisante, puisqu'elle a été démontrée à partir du cas particulier d'un univers constitué d'une seule source lumineuse ponctuelle.

#### 3.4.9.2 Mapping euclidien

Dans tous les dispositifs développés jusqu'à aujourd'hui, la mise en correspondance utilise des fonctions de mapping monotones basée sur une représentation euclidienne de l'espace de l'image (fig. 3.46), où la position des champs récepteurs est exprimées à partir de ses coordonnées cartésiennes  $\{(x_i, y_i), i \in [1..N]\}$  (Fish, 1976, Meijer, 1992, Capelle *et al.*, 1998, Cronly-Dillon, 2001).

Dans ce cas, l'analogie est basée sur les propriétés d'interprétation "naturelles" du système

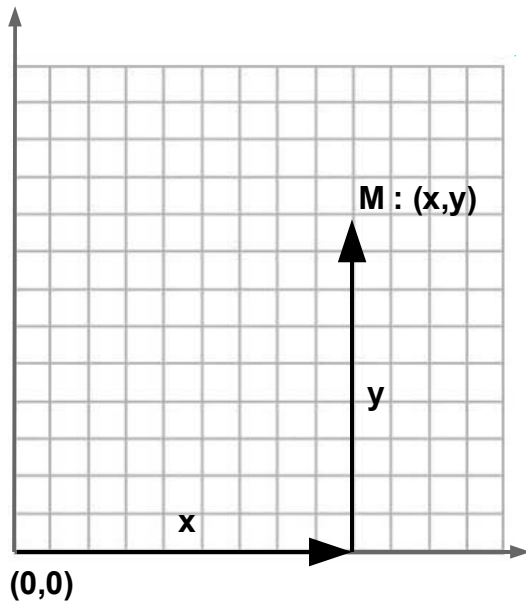


Illustration 3.46 :représentation cartésienne de l'image. La notion de voisinage est définie en terme de translations

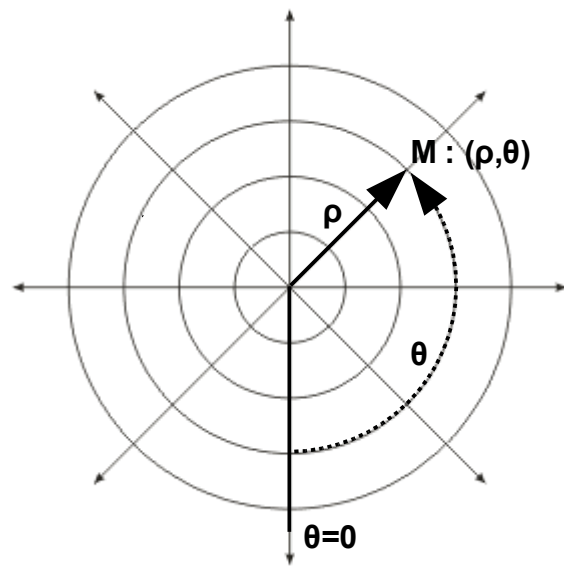


Illustration 47 : représentation polaire de l'image. La notion de voisinage est définie en terme de zooms et de rotations.

auditif. En particulier, il a été montré que la hauteur apparente d'un son pur dépendait de sa fréquence. Un son aiguë paraîtra plus haut (spatialement) qu'un son grave (Roffler et Buttler, 1968, cité par Fish, 1976). La tonie du signal semble donc une échelle naturelle pour correspondre à la position verticale de la source sonore. Par ailleurs, l'azimut de la source semble naturellement indiquée pour transcrire sa position horizontale dans l'image.

Selon les dispositifs, la position horizontale est parfois renforcée par un balayage temporel (Fish, 1976, Meijer, 1992) ou par une nouvelle composante fréquentielle (Capelle, 1998). Pour notre part, nous avons choisi de n'utiliser aucun renforcement concernant la position horizontale. En particulier, l'utilisation d'un balayage temporel, en limitant le taux de rafraichissement de l'image, est susceptible de perturber l'établissement d'un couplage sensori-moteur. Nous nous sommes par ailleurs interrogés sur le fait d'utiliser une composante fréquentielle pour coder la position horizontale. Nous avons toutefois considéré que cela pourrait réduire le couplage sensorimoteur en ce sens que les variations fréquentielles sont associées à tous les mouvements et non plus seulement aux mouvements selon l'axe vertical. La discussion à ce sujet dans Capelle *et al.* est toutefois très intéressante et ce type de codage mériterait d'être testé avec TheVIBE.

Dans le cas du mapping euclidien, nous avons donc associé à chaque champ récepteur une source sonore dont l'azimut et la tonie dépendent respectivement de sa position horizontale et verticale selon une loi linéaire (donc strictement monotone). Les paramètres du signal (ILD, fréquence) sont calculés ensuite en fonction de l'échelle considérée (voir §3.3.2 et §3.3.3).

En pratique :

Soit H et W la hauteur et la largeur de l'image issue de la caméra (en pixels).

Soit  $(x_i, y_i)$  les coordonnées du centre du champ récepteur  $i$  dans le référentiel de l'image centré sur le bord en haut à gauche,

Le paramètre d'azimut s'écrit :  $a_i = \frac{x_i}{W}$

Le paramètre de tonie s'écrit :  $t_i = (1 - \frac{y_i}{H}) \times (t_{max} - t_{min}) + t_{min}$

où tmin et tmax sont les bornes respectives des valeurs de tonies dans l'échelle considérée.

### 3.4.9.3 Mapping polaire

Dans ce cas, on considère la position du champ récepteur exprimé dans un repère polaire. Nous associons alors au champ récepteur une source sonore dont la tonie et l'azimut dépendent respectivement de la norme  $\rho$  et de l'argument  $\theta$  du centre du champ récepteur (fig. 3.47).

En pratique, nous avons choisi pour  $\rho$  une fonction de mapping décroissante : plus  $\rho$  est grand (plus le champ récepteur est en périphérie), plus la tonie du signal est basse. L'argument pour cela est que la résolution du système auditif reste plus fine en haute qu'en basse fréquence (voir fig. 3.47), même si l'utilisation d'une échelle logarithmique ou d'une échelle en Bark réduit ces différences. Nous choisissons donc de privilégier les capacités de discrimination entre les sources sonores au centre de l'image.

Pour l'azimut, nous avons utilisé une fonction linéaire décroissante de  $\theta$ . Pour  $\theta=0$ ,  $a=1$  (position apparente à l'extrême droite) et  $a=0$  pour  $\theta=2\pi$  (position apparente à gauche). On notera cette fois la présence d'une discontinuité en  $\theta=2\pi$ , puisque le paramètre "a" passe directement de 0 à 1 (fig. 3.28). Il est difficile de savoir si cette discontinuité va perturber concrètement l'apprentissage du couplage sensori-moteur. Le cas échéant, le codage sur un demi cercle seulement est envisageable.

#### En pratique :

Dans le cas d'un mapping polaire, on ne s'intéresse qu'à une fenêtre circulaire de l'image, en particulier en lien avec les effets de bords induits par un échantillonnage logarithmique destiné à être utilisé conjointement (nous reviendrons la-dessus au prochain paragraphe).

On considère donc  $H$  la taille du plus petit côté de l'image qui constitue le diamètre de la fenêtre de travail.

Le paramètre d'azimut s'écrit :

$$a_i = 1 - \frac{\theta_i}{2\pi}$$

Le paramètre de tonie s'écrit :

$$t_i = (1 - \frac{\rho_i}{r}) \times (t_{max} - t_{min}) + t_{min}$$

avec  $r = H/2$  le rayon de la fenêtre de travail.

### 3.4.9.4 Mapping log-polaire

Le terme de mapping log-polaire que nous utiliserons par la suite est un abus de langage. Il s'agit en fait du couplage d'une fonction de mapping polaire (définie au paragraphe précédent) et d'un échantillonnage spatialement variant de type "logarithmique" défini au paragraphe 2.5.



Il est impossible de faire une analogie formelle entre le mapping log-polaire observé au niveau du cortex visuel et le mapping log-polaire telle que nous le définissons ici. En effet, il faudrait pour cela démontrer que le cortex auditif est organisé géométriquement selon un espace cartésien dont un axe serait l'azimut et l'autre la tonie, ce qui est peu vraisemblable.

En revanche, il existe véritablement un rapport de type projection log-polaire entre un mapping cartésien croisé avec un échantillonnage uniforme de l'image et un mapping polaire croisé avec un échantillonnage spatialement variant : formellement le fonctionnement de TheVIBE configuré avec un mapping log-polaire est analogue au fonctionnement du dispositif TheVIBE configuré avec un mapping cartésien-uniforme utilisant en entrée une image ayant subie une transformation log-polaire.

## Références

- Auvray, M.; Hanneton, S.; Lenay, C. & O'Regan, K. (2005), 'There is something out there: distal attribution in sensory substitution, twenty years later', *Journal of Integrative Neuroscience* **4**(4), 505.
- Blauert, J. (1997), *Spatial Hearing*, The MIT Press.
- Bonnet, C. (1986), *Manuel pratique de psychophysique*. Armand Collin
- Bouquet J-Y. (2008) Matlab Calibration Toolbox. [www.vision.caltech.edu/bouquetj/calib\\_doc](http://www.vision.caltech.edu/bouquetj/calib_doc).
- Buser P. et Imbert M. (1987) Audition. Hermann, Paris.
- Capelle, C.; Trullemans, C.; Arno, P. & Veraart, C. (1998), 'A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution', *Biomedical Engineering, IEEE Transactions on* **45**(10), 1279—1293.
- Chen, S.; Lovell, N. & Suaning, G. (2004), 'Effect on prosthetic vision visual acuity by filtering schemes, filter cut-off frequency and phosphene matrix: a virtual reality simulation.', *Conf Proc IEEE Eng Med Biol Soc* **6**, 4201--4204.
- Cronly-Dillon, J.; Persaud, K. & Gregory, R. P. (1999), 'The perception of visual images encoded in musical form: a study in cross-modality information transfer.', *Proc Biol Sci* **266**(1436), 2427—2433.
- Duchêne C. (2007) Influences des fluctuations aléatoires sur l'échantillonnage et la quantification du système visuel. Thèse de Doctorat. Université Joseph Fourier, Grenoble.
- Fish, R. M. (1976), 'An audio display for the blind.', *IEEE Trans Biomed Eng* **23**(2), 144--154.
- Garner, W. R. (1953), 'An informational analysis of absolute judgments of loudness.', *J Exp Psychol* **46**(5), 373--380.
- Garner, W. R. & Hake, H. W. (1951), 'The amount of information in absolute judgements.', *Psychol Rev* **58**(6), 446--459.

- Haftner, E. R.; Dye, R. H.; Nuetzel, J. M. & Aronow, H. (1977), 'Difference thresholds for interaural intensity.', *J Acoust Soc Am* **61**(3), 829--834.
- Hartmann, W. M. Springer-Verlag, ed. (1997), *Signals, Sound, and Sensation*.
- Hecht, S., and Verrijp, C. D. (1933). Intermittent stimulation by light. II. The relation between intensity and critical fusion frequency for different retinal locations. *J gen Physiol* **17**, 251-265.
- Hershkowitz, R. M. & Durlach, N. I. (1969), 'Interaural time and amplitude jnds for a 500-Hz tone.', *J Acoust Soc Am* **46**(6), 1464--1467.
- Hérault, J. & Jutten, C. (2001), 'Réseaux neuronaux et traitement du signal, manuel du DEA de Sciences Cognitives.', Institut National Polytechnique de Grenoble, Grenoble.
- Kaczmarek, K. & Bach-y-Rita, P. (1995), *Advanced Interface Design and Virtual Environments*, Oxford University Press, chapter Tactile Displays, pp. 349-414.
- Melchior et Menvielle (2009) Cours de statistique Inserm. <http://ifr69.vjf.inserm.fr/~u88/site/Cours%20sondages%202005.pdf>
- McCroskey, R. & Keith, R. (1996), 'Auditory fusion test - revised : Instruction and User's manual', Auditec of St Louis: St Louis MO.
- Meijer, P. B. (1992), 'An experimental system for auditory image representations.', *IEEE Trans Biomed Eng* **39**(2), 112--121.
- Miller, G. A. (1956), 'The magical number seven plus or minus two: some limits on our capacity for processing information.', *Psychol Rev* **63**(2), 81--97.
- Miller, G. A. (1948), 'The Perception of Short Bursts of Noise', *J. Acoust. Soc. Amer.* **20**, 160-170.
- Rayleigh (1907), 'On our perception of sound direction', *Philosophical magazine* **13**, 214-232.
- Shannon, C. & Weaver, W. of Illinois Press, U., ed. (1949), *The Mathematical Theory of Communication*, Urbana.
- Viemeister, N. F. (1979), 'Temporal modulation transfer functions based upon modulation thresholds.', *J Acoust Soc Am* **66**(5), 1364--1380.
- Zwicker, E. (1961), 'Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)', *The Journal of the Acoustical Society of America* **33**, 248.
- Zwicker, E.; Floitrop, G. & Stevens, S. (1957), 'Critical Band Width in Loudness Summation', *Journal of the acoustical society of america* **29**, 548-557.



## Chapitre 4 : Premiers pas avec le dispositif TheVIBE

Maintenant que le dispositif de substitution est opérationnel et configuré et que nous disposons d'une simulation temps réel du traitement du signal dans le système visuel primaire, il s'agit d'étudier si, en pratique, le dispositif est fonctionnel et de voir dans quelle mesure l'introduction de cette simulation dans la conception du dispositif permet en pratique d'en améliorer l'efficacité. Pour cela, nous devons donc mettre au point des méthodes d'évaluation qui nous permettent d'opérationnaliser et de quantifier l'efficacité du dispositif et de tester l'intérêt des traitements que nous proposons.

Le dispositif TheVIBE, parce qu'il est récent, a été très peu testé. Au début de cette étude, la seule expérience publiée à son sujet était celle de Malika Auvray (2005).

Dans cette expérience, les participants étaient amenés à manipuler le dispositif dans des conditions volontairement restreintes (par exemple, ils ne pouvaient pas bouger la tête, devaient manipuler l'objet avec un levier, etc...). Ils devaient par la suite choisir un scénario décrivant le dispositif parmi plusieurs (par exemple "le dispositif produit des sons en fonction de capteurs placés sur mon bras", ou encore "le dispositif produit des sons aléatoires qui ne sont pas en lien avec mes propres actions", etc...). Ce type de protocole, inspiré de Epstein (1986), permet d'étudier comment les conditions d'utilisation conditionnent le sens que les sujets attribuent au dispositif, et plus particulièrement ici, permettent ou non l'établissement d'une expérience distale.

La question de savoir si le dispositif TheVIBE pouvait être d'une quelconque utilité en pratique restait donc entière. C'est à cette question que nous nous sommes attelés. Toutefois, plutôt que de nous lancer directement dans des tests impliquant des tâches complexes, nous avons choisi de tester graduellement les possibilités du dispositif. Pour cela nous nous sommes inspirés du développement de la vision chez l'enfant. Le lecteur intéressé pourra se reporter à l'ouvrage de Michèle Mazeau (1995).

À la naissance, le système visuel du nouveau-né n'est pas opérationnel. À vrai dire, il n'est considéré comme "mature" qu'à partir de l'âge de 13 ans. Les étapes les plus significatives de son développement ont toutefois lieu dans les premiers mois de la vie. Les 6 premiers mois en particulier, on observe "un développement rapide des fonctions visuelles en parallèle et en interaction avec le développement des mouvements oculaires" (FNRO, 2009).

Les mouvements oculaires en particulier apparaissent dès la naissance et sont considérés comme l'un des actes moteurs les plus précoces. Les mouvements de saccades vers des objets d'intérêt et les mouvements oculaires pour suivre une cible sont, dès l'âge de 4 mois, aussi rapide que chez l'adulte. C'est vraisemblablement l'une des premières compétences visuelles du nouveau-né.

Selon Zahlongi (2001) : "Les mouvements des globes oculaires, leurs orientations, ont pour but d'amener le stimulus juste sur la fovéa. Cette saisie fovéale est la condition première de toute appropriation du monde environnant". C'est sur cette capacité de "saisie fovéale" que nous nous sommes penchés en premier lieu, et que nous développerons dans la partie 1.

Une autre capacité précoce associée à la vision est la coordination vision-préhension, opérationnelle dès 4 mois. Nous proposons donc dans un deuxième temps une expérience de localisation-préhension, où le sujet doit chercher une cible lumineuse et s'en emparer. Cette expérience est décrite dans la deuxième partie du chapitre.

## 4.1 Saccades vers une cible à l'aide du dispositif TheVIBE

Dans cette première expérience, nous avons voulu tester les capacités des sujets à pointer le dispositif vers une cible lumineuse. Cette tâche s'inspire directement des tâches de saccade oculaire.

Définir en quoi consiste une saccade oculaire semble à priori évident : il s'agit de "diriger son regard" vers une cible donnée. Mais que signifie au juste "diriger son regard" ? Une expression identique souvent donnée en consigne aux sujets d'une expérience d'oculométrie est de "fixer son regard" sur la cible. Si l'on prend l'expression au sens littéral, il suffit pour réaliser la tâche que la cible soit dans mon champ de vision et que je la maintienne fixe dans ce champ de vision. Un sujet qui opérerait de la sorte serait vraisemblablement écarté des analyses.

Ce propos, volontairement naïf, a pour but d'attirer l'attention sur la question de la définition du "regard" et de la notion de "fixation". Implicitement, ces deux notions font référence à une zone particulière du champ visuel, que l'on pourrait qualifier de "centre du champ visuel". On considère généralement que cette zone correspond à la fovéa, qui s'étend sur environ 3 ° de champ visuel autour de l'axe optique de l'oeil, et dans laquelle la densité des photorécepteurs, et donc la résolution de l'oeil, est la meilleure (voir chap 2). D'où l'expression "saisie fovéale" dans l'introduction de ce chapitre.

Ce point de vue est implicite dans les expériences d'oculométrie, de sorte que les sujets comprennent instantanément ce qu'il faut faire lorsqu'on leur demande de fixer une cible : chez un sujet voyant, la capacité à placer un objet au niveau de la fovéa est si automatique qu'elle paraît aller de soi.

Pour l'utilisateur débutant d'un dispositif de substitution, ce n'est pas le cas, et ce pour deux raisons :

1. La capacité à fixer son "regard" (qu'il s'agisse de bouger un oeil ou une caméra) n'est pas immédiate et nécessite un apprentissage (rappelons que cet apprentissage pour le système visuel humain dure 4 mois).
2. Seuls certains dispositifs présentent un échantillonnage spatialement variant et donc une fovéa comme par exemple le PSVA. Il n'y a donc pas toujours de "fovéa" qui puisse définir un "centre".

Dans cette expérience, nous avons donc étudié comment les sujets réalisaient une tâche de fixation avec le dispositif TheVIBE. Nous avons en particulier testé l'effet d'un échantillonnage spatialement variant biomimétique présentant une "fovéa" sur le comportement des sujets dans cette tâche.

### 4.1.1 Méthode

Nous avons pour cela mis en place un dispositif expérimental permettant d'afficher, sur un grand écran, une cible visuelle blanche sur un fond noir (fig. 4.1).

Ce dispositif est constitué d'un vidéo-projecteur placé à trois mètres derrière un écran translucide. Les sujets équipés du dispositif TheVIBE sont placés face à l'écran de sorte que le champ de la caméra embrasse la totalité de la zone de projection, et qu'ainsi toutes les cibles apparaissent dans le champ de la caméra.

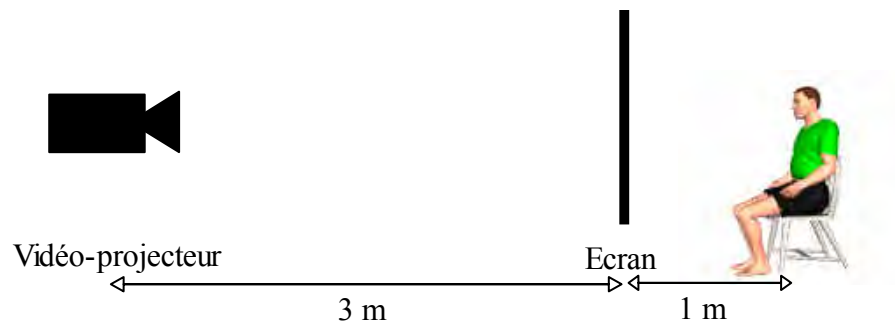


Figure 4.1 : le dispositif expérimental. L'image affichée par le video-projecteur correspond exactement à la taille du champ de la caméra

#### 4.1.1.1 Appareillage et dispositif expérimental

Les sujets portent un masque semi-rigide (du type de ceux distribués dans les avions) doublé de disques en coton pour garantir une occultation complète et efficace du champ visuel. Ils sont équipés du dispositif TheVIBE comprenant :

- Une webcam Quikcam Pro 5000 présentant un champ de 90° en diagonale fixée à l'arceau du casque juste au dessus de leur tête.
- Un casque stéréophonique Sennheiser HD 280 Pro dont la réponse en fréquence est particulièrement fidèle
- Un ordinateur de bureau Dell Optiplex GX 620 équipé d'un processeur Intel cadencé à 3 Ghz, et de 2 Go de mémoire vive.

L'image de la webcam est traitée par le logiciel TheVIBE qui génère un son transmis au sujet par les écouteurs.

##### **Version du logiciel :**

Pour cette expérience, le PC est équipé de la version 2.0 de TheVIBE. Cette nouvelle version du logiciel à été réalisée spécialement pour cette expérience par Sylvain Hanneton.

Cette version possède notamment la possibilité de configurer le logiciel dans des fichiers externes, le fichier d'option et le fichier de configuration, et une fonction permettant d'enregistrer en temps réel l'activité des champs récepteurs. Comme nous le verrons, cette dernière fonction va nous permettre de retrouver la position de la cible dans l'image de la caméra en fonction du temps.

##### **Méthode d'échantillonnage :**

Le fichier de configuration, appelé "fichier de rétine", unique pour chaque utilisateur, comprend 192 champs récepteurs dont les centres sont répartis par division de l'image en rectangles élémentaires (ch.3 §2.3). L'image est découpée en 16 x 12 rectangles de 20 pixels de côté. Un point est choisi dans chaque rectangle et constitue le centre d'un champ récepteur. Pour chaque champ récepteur, on choisit 10 échantillons dans un carré de 20 pixels de côté autour de chaque centre.

Pour les sujets du groupe "Log2" et du groupe "log4", une transformation supplémentaire est appliquée sur les échantillons de la rétine pour simuler l'échantillonnage spatialement variant au niveau des photorécepteurs (voir chap. 2 et chap. 3). Pour le groupe "Log2", le paramètre R0 de magnification centrale est fixé à 2, pour les sujets du groupe "Log4", ce paramètre est fixé à 4. Les

fichiers de rétine du groupe "Unif" ne subissent aucune modification.

La figure 4.2 illustre graphiquement la distribution des champs récepteurs et des échantillons dans l'image. Les échantillons d'un même champ récepteur sont rattachés à leur centre par un trait noir. On remarquera d'ors et déjà que l'application de la transformation spatialement variante produit des zones non-échantillonnées sur les bords de l'image. Cet effet est inévitable lorsque l'on travaille sur une fenêtre carrée. A l'époque nous n'avions pas d'alternatives pour corriger ce défaut. Nous reviendrons sur ce point dans la discussion.

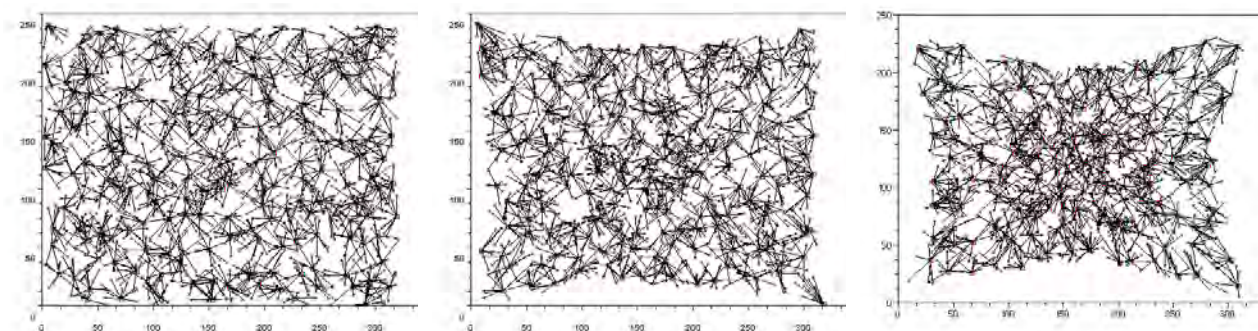


Figure 4.2 : "Fichiers de rétine" pour l'expérience sur la localisation. De gauche à droite, rétines des groupes "Unif", "Log2" et "Log4" respectivement. Les carrés sont les centres des champs récepteurs et les croix les échantillons.

### Mapping visuo-auditif :

Le mapping visuo-auditif est de type "Euclidien" (ch. 3 §3.5.1) pour tous les sujets. Par ailleurs :

- Le codage de l'activité des champs récepteurs utilise une échelle linéaire d'amplitude (voir Ch. 3 §3.3.1)
- Le codage du panoramique utilise également une différence d'amplitude linéaire entre les deux oreilles (voir Ch. 3 §3.3.2)
- Le codage de la tonie se fait selon une échelle linéaire en fréquence, dans l'intervalle 300-3000 Hz (voir Ch. 3 §3.3.3).

Il est important de noter que l'attribution des paramètres sonores des sources associées aux champs récepteurs est faite à partir de la position du centre des champs récepteurs avant une éventuelle modification de la position des échantillons pour réaliser un échantillonnage spatialement variant sans changer l'aptitude à discriminer les sources sonores entre elles.

### Environnement :

Les sujets sont placés face à un écran dans une pièce sombre. Un vidéo-projecteur projette sur l'écran l'image d'un disque blanc sur un fond noir. La taille de l'image projetée sur l'écran correspond à la taille du champ de la caméra. Le centre du vidéo projecteur et le centre de l'image sont alignés avec le centre optique de la caméra, de manière à ce que toute cible apparaisse dans le champ du dispositif. Un premier contrôle visuel de l'alignement est effectué en début d'expérience et corrigé si nécessaire en modifiant la position et la hauteur du siège du sujet.

Les cibles lumineuses sont générées à l'aide de la psychtoolbox (Brainard, 1997) fonctionnant avec le logiciel Matlab. Les cibles correspondent à des disques blanc dont le diamètre est d'environ 8° dans le champ de vision de la caméra.





Figure 4.3 : dispositif expérimental. Durant la phase d'apprentissage libre, le sujet est debout face à la cible



Figure 4.4 : dispositif expérimental. Durant les phases d'apprentissage et de test, le sujet est assis face à l'écran.

Un écran de contrôle affiche en permanence l'image issue de la caméra (figures 4.3 et 4.4 , devant l'expérimentateur).

#### 4.1.1.2 Protocole expérimental

Le protocole expérimental est très largement inspiré des travaux de Malika Auvray (Auvray, 2004).

En début de séance, on explique au sujet le but du dispositif et son principe de fonctionnement. On explique en particulier le principe de codage du dispositif : le lien entre la hauteur dans l'image et la hauteur tonale du son, et entre la position latérale dans l'image et le panoramique du son. Le sujet est ensuite équipé avec le dispositif.

##### **Réglage de l'alignement du dispositif :**

On fait assoir le sujet face à l'écran sur lequel on affiche un carré blanc matérialisant le centre de l'image. On règle alors la position et la hauteur du siège de manière à ce que la caméra soit alignée avec le vidéo-projecteur et le centre de l'image à l'écran.

On demande ensuite au sujet de se mettre au repos, c'est à dire comme s'il regardait droit devant lui. A l'aide de l'écran de contrôle, on effectue alors un réglage fin de la position de la caméra (qui peut pivoter sur son socle) de manière à ce que le champ de la caméra recouvre la totalité de l'image. On demande au sujet de mémoriser cette posture qu'il sera amené à reprendre entre chaque essai.

L'expérience est ensuite divisée en trois phases séparées par des pauses de 5 minutes : la familiarisation libre, l'apprentissage du dispositif, et enfin le test.

##### **La familiarisation libre**

Durant la phase de familiarisation libre, le sujet est debout à environ 50 cm de l'écran (fig 4.3). Une cible lumineuse est affichée en face de lui.



Dans un premier temps, on demande au sujet d'observer cette cible comme il le veut, en l'engageant à bouger la tête et à explorer les limites du dispositif. Cette première phase dure 5 minutes.

On demande ensuite au sujet de faire des pas sur la gauche et sur la droite tout en maintenant la cible dans le champ de la caméra. Après quelques passages, on recule le sujet à environ 1 m de l'écran et on lui demande de recommencer la même tâche. Cette deuxième phase dure 5 minutes également. Cette tâche se veut l'écho de l'étape de "contact" décrite par Malika Auvray, qui correspond à l'apprentissage des lois de covariation sensorimotrices permettant de maintenir une stimulation constante. L'apprentissage du "contact" constitue selon Malika la première étape dans l'immersion d'un dispositif de substitution.

### **L'apprentissage du dispositif**

La phase de prise en main est basée sur un exercice de masquage où le sujet doit masquer la cible à l'aide de sa main. Selon Auvray (2005), ce type de masquage (effectué à l'aide d'un carton rigide dans l'article en question) favorise "l'attribution distale" qui constitue la deuxième étape d'immersion dans un dispositif de substitution. Cette étape correspond à "la compréhension, par les utilisateurs, que l'origine de leurs sensations organiques est due à la rencontre avec un objet provenant de l'espace perceptif ouvert par l'outil" (Auvray, 2004, p 87).

Durant la phase d'apprentissage, le sujet est assis. Au début de chaque essai, il doit se remettre dans la "position de repos" préalablement déterminée. Un petit carré blanc apparaît au centre de l'écran pour permettre à l'opérateur de contrôler la position du sujet via l'écran de contrôle. Si le carré n'est pas au centre, l'opérateur fournit des indications verbales pour ramener le carré au centre de l'image de la caméra. Le champ de la caméra embrasse alors l'intégralité de l'image projetée. L'opérateur démarre l'essai lorsque la position du sujet est correcte.

Un cible est alors affichée à une position aléatoire dans l'image. Le sujet doit, en gardant le bras tendu, masquer cette cible. L'opérateur valide l'essai lorsque la cible est effectivement masquée par la main (et non lorsque la cible sort du champ de la caméra, ce qui pour le sujet correspond à ne plus entendre la cible, tout comme si elle était masquée). Le fait d'utiliser des positions aléatoires pour la cible permet d'empêcher que le sujet apprenne par coeur les positions.

Durant les trente premiers essais, on demande au sujet de décomposer le masquage en deux phases. Il doit d'abord bouger sa tête afin de "pointer son dispositif vers la cible, c'est à dire mettre la cible au centre de son champ perceptif". La définition de la notion du "centre du champ perceptif" est laissée au soin du sujet, nous reviendrons sur ce point dans la discussion. Il doit ensuite placer sa main de manière à masquer la cible, c'est-à-dire dans l'axe de visée de la caméra.

Durant les trente essais suivants, on demande au sujet de masquer la cible directement, c'est à dire sans bouger la tête.

La durée mise par le sujet pour réaliser les soixante essais est enregistrée. C'est ce que l'on a appelé le "temps d'apprentissage" (TA).

### **Le test**

Durant la phase de test, le sujet est assis en face de l'écran. Il est informé qu'il s'agit de la phase de test. Chaque essai démarre par un contrôle de la position du sujet, qui doit se remettre en "position de repos".

Une fois la position validée par l'opérateur, une cible est affichée à l'écran. La position de la cible correspond à l'une des 20 positions standards (fig. 4.5). Le sujet doit alors "pointer son dispositif

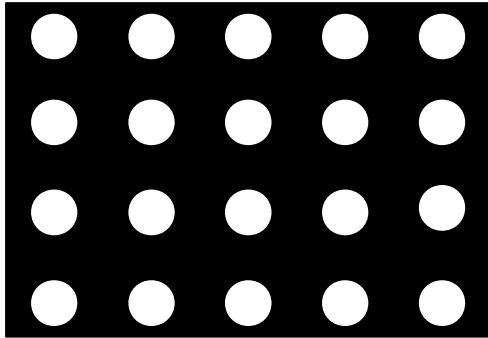


Figure 4.5 : phase de test. Les 20 positions standards de la cible dans l'image

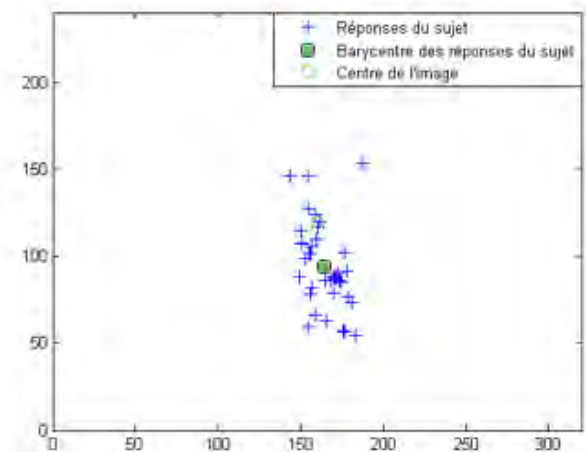


Figure 4.6 : positions de la cible dans l'image au moment où le sujet a validé chaque essai.

vers la cible, c'est à dire mettre la cible au centre de son champ perceptif". Il valide l'essai lorsqu'il pense que c'est le cas.

La phase de test est constituée de 40 essais, chaque position possible de la cible apparaissant deux fois. La durée entre l'affichage de la cible et le moment où le sujet valide l'essai est enregistré : c'est ce que nous avons appelé le "temps de réponse" du sujet. L'activité des champs récepteurs est également enregistrée une fois toute les 5 images (soit 5 fois par secondes). Ceci nous permet de suivre en temps réel la position de la cible dans l'image et en particulier d'enregistrer la position de la cible dans l'image à la fin de l'essai. Pour cela, on calcule le barycentre des centres des champs récepteurs pondérés par la moyenne de leur activité sur les trois dernières mesures avant la validation. La figure 4.6 présente les mesures de la position de la cible à la fin des 40 essais d'un sujet (croix).

A l'issu du test, on calcule le barycentre des positions de la cible à la fin des 40 essais. C'est ce que nous appelons le "centre subjectif" du sujet (fig. 4.6, cercle plein). Nous calculons enfin la distance moyenne entre la position de la cible à la fin de chaque essai et le "centre subjectif" du sujet, ce qui définit notre mesure de "Précision".

#### 4.1.2 Participants et plan d'expérience

Vingt et un sujets, 15 hommes et 6 femmes, voyants et sans problèmes auditifs avérés ont pris part à l'expérience. L'âge moyen des sujets est de 26 ans avec un écart-type de 4 ans. Aucun d'entre eux n'avait utilisé ce type d'appareil auparavant. Les sujets ont été répartis dans trois groupes (Unif, Log2, Log4) appariés pour leur sexe et leur âge.

Le plan expérimental est donc entièrement inter :  $S_{21} < E_3 >$

où le facteur indépendant E correspond au type d'échantillonnage de l'image dans le dispositif: uniforme ("Unif"), spatialement variant avec un taux de magnification de 2 ("Log2"), et spatialement variant avec un taux de magnification de 4 ("Log4").

### 4.1.3 Résultats

Le tableau 1 présente le temps d'apprentissage ainsi que le temps de réponse moyen et la précision mesurés lors du test pour chaque sujet, ainsi que leur groupe respectif.

Tableau 1 : données de l'expérience de localisation

Sujet	Groupe	TA (min)	TR (sec)	Pr (pixels)
1	Unif	46	24,68	32,48
2	Unif	37	37,36	48,87
3	Unif	36	24,02	62,86
4	Unif	28	8,85	54
5	Unif	32	21,24	46,19
6	Unif	35	22,59	32,24
7	Unif	40	11,47	37,95
8	Log2	23	13,13	8,23
9	Log2	20	9,55	32,52
10	Log2	22	7,26	56,84
11	Log2	32	12,36	49,41
12	Log2	21	6,05	38,59
13	Log2	23	11,39	10,61
14	Log2	23	14,29	46,99
15	Log4	15	9,22	21,17
16	Log4	18	9,89	22,36
17	Log4	19	7,89	15,56
18	Log4	26	13,85	21,98
19	Log4	19	9,63	27,78

#### 4.1.3.1 Analyse du temps d'apprentissage

La figure 4.7 présente les moyennes des temps d'apprentissage en fonction des groupes.

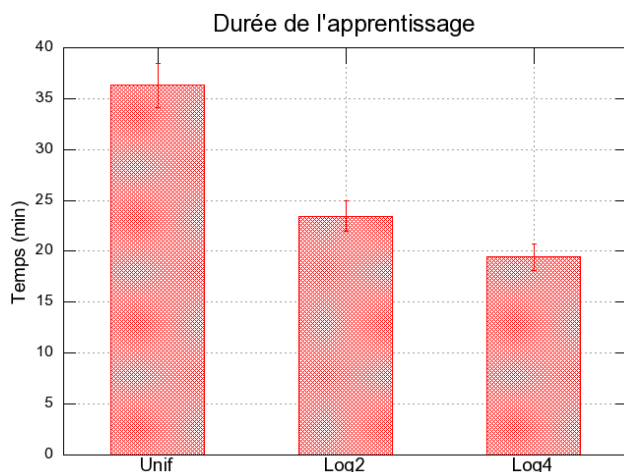


Figure 4.7 : temps d'apprentissage moyen par groupe

Pour déterminer le test statistique approprié pour comparer les temps d'apprentissage, nous avons testé la normalité des distributions de temps d'apprentissage par groupe à l'aide d'un test de Shapiro-Wilk (1965). La distribution des données dans le groupe Log2 apparaît significativement différente d'une distribution normale (Test de Shapiro-Wilk,  $p < 0.01$ ). Nous avons dans un premier temps tenté d'appliquer une transformation logarithmique sur les données. La distribution du logarithme

des temps d'apprentissage dans le groupe "Log2" reste toutefois significativement différente d'une distribution normale ( $p < 0.05$ ). Nous avons donc opté pour un test non-paramétrique : le test de Wilcoxon (1945), avec lequel nous avons comparé les temps d'apprentissage du groupe "Unif" avec ceux des groupes "Log2" et "Log4" (Contraste C1) puis les temps d'apprentissage du groupe "Log2" avec ceux du groupe "Log4" (Contraste C2). En l'absence d'hypothèses théoriques sur le sens de l'effet, nous avons effectué des tests bilatéraux.

Les deux contrastes sont significatifs (Test de Wilcoxon bilatéral,  $p < 0.001$  pour C1 et  $p < 0.05$  pour C2). Le temps d'apprentissage est donc significativement plus court dans les groupes avec un échantillonnage spatialement variant ("Log2" et "Log4") que dans le groupe avec échantillonnage uniforme ("Unif"), et significativement plus court dans le groupe "Log4" que dans le groupe "Log2".

#### 4.1.3.2 Analyse du test

La figure 4.8 présente les résultats du test : temps de réponse moyen et précision des sujets en fonction de leur groupe.

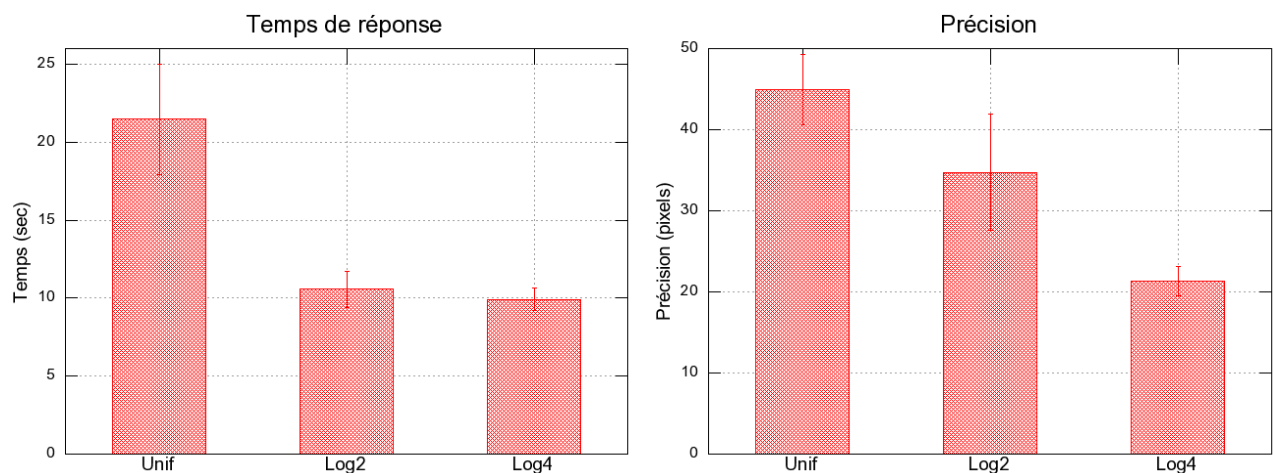


Figure 4.8 : performances moyenne des sujets lors du test de fovéalisation, en fonction de leur groupe

Aucune distribution n'est apparue significativement différente d'une distribution normale (Test de Shapiro-Wilk,  $p > 0.05$  pour toutes les distributions). Nous avons donc utilisé le test de Student avec lequel nous avons testé les mêmes contrastes que dans le cas de l'apprentissage :

C1 = groupe "Unif" versus groupe "Log2" et "Log4"

C2 = groupe "Log2" versus groupe "Log4"

En l'absence d'hypothèses théoriques sur le sens de l'effet, nous avons effectué des tests bilatéraux. Les résultats sont consignés dans le tableau 2.

Tableau 2 : tests statistique des contrastes sur les temps de réponse et sur la précision des sujets

Contraste	Temps de réponse	Précision
C1 : "Unif" vs "Log2, Log4"	Significatif ( $p < 0.05$ )	Significatif ( $p < 0.05$ )
C2 : "Log2" vs "Log4"	Non significatif ( $p = 0.71$ )	Non significatif ( $p = 0.21$ )

Les sujets utilisant une "rétine" spatialement variante (Log2, Log4) ont donc un temps de réponse significativement plus court que les sujets utilisant une "rétine" uniforme (Unif). En revanche, on ne trouve pas de différence significative de performance entre les sujets du groupe "Log2" et ceux du groupe "Log4".

## 4.1.4 Discussion

Dans cette expérience, nous avons testé l'effet d'un échantillonnage spatialement variant dans le dispositif TheVIBE sur les performances de sujets dans une tâche de "saccade" vers des cibles lumineuses.

Nous avons pour cela comparé les performances des sujets pour lesquels l'échantillonnage de l'image dans le dispositif était uniforme aux performances des sujets pour lesquels l'échantillonnage était spatialement variant. Par ailleurs, nous avons testé, chez les sujets utilisant un échantillonnage spatialement variant, l'effet du taux de magnification de l'échantillonnage.

### 4.1.4.1 Une tâche de saccade?

Le lecteur s'interrogera peut-être sur l'équivalence de notre protocole de test avec un protocole d'oculométrie standard. Notre tâche peut paraître en particulier moins contrainte qu'une tâche de fixation visuelle traditionnelle, à l'aide d'un eye-tracker par exemple, puisque le centre de fixation est défini *a posteriori* pour chaque sujet, en calculant la moyenne de ses réponses.

Il n'en est rien : en effet, dans une tâche de saccade visuelle, la position de l'oeil est préalablement calibrée en demandant au sujet de "fixer", justement, une cible au centre de l'écran. Toutes les positions de l'oeil sont ensuite rapportées à cette position initiale. Rien ne dit pour autant qu'au moment de la calibration la cible était véritablement en coïncidence avec le centre "objectif" de la rétine, qu'il serait d'ailleurs délicat de définir. La seule différence, dans notre cas, est que la calibration, au lieu d'être réalisée une fois sur une cible arbitraire au début de l'expérience, est réalisée en calculant *a posteriori* le barycentre de la position de toutes les cibles fovéalisées.

### 4.1.4.2 Principaux résultats

Dans cette expérience, nous mettons effectivement en évidence un effet du mode d'échantillonnage sur le comportement des sujets.

Les performances des sujets utilisant un échantillonnage spatialement variant (groupes "Log2" et "Log4") sont significativement meilleures que celles des sujets utilisant un échantillonnage uniforme. Le temps de réponse est pratiquement réduit de moitié :  $10.25 \pm 0.6$  sec dans les groupes "spatialement variants" contre  $21.5 \pm 3.55$  sec dans le groupe "uniforme". La précision passe quant à elle de  $45 \pm 5$  pixels (soit environ  $11^\circ$  de champ visuel) dans le groupe "uniforme" à  $28 \pm 4$  pixels (soit environ  $7^\circ$  de champs visuel) dans les groupes "spatialement variants".

Par ailleurs, le temps mis pour réaliser l'exercice d'apprentissage est également significativement plus court dans le cas d'un échantillonnage spatialement variant que dans le cas d'un échantillonnage uniforme. Dans les groupes "spatialement variants", le temps mis pour réaliser l'exercice d'apprentissage est de  $21 \pm 1$  minutes en moyenne, contre  $36 \pm 5$  minutes dans le groupe "uniforme".

L'effet du taux de magnification de l'échantillonnage spatialement variant, qui différencie les groupes "Log2" et "log4", est lui beaucoup moins net. Nous n'observons son effet que sur les durées

de l'exercice d'apprentissage : les sujets du groupe "Log4" mettent en moyenne  $19 \pm 3$  minutes , contre  $23 \pm 4$  minutes dans le groupe "Log2".

L'implémentation d'un échantillonnage spatialement variant dans le dispositif a donc un effet très marqué sur les performances des sujets. Avec un temps d'apprentissage réduit presque de moitié, les sujets bénéficiant d'un échantillonnage spatialement variant sont deux fois plus rapides et plus précis qu'avec un échantillonnage uniforme.

#### 4.1.4.3 Analyse des résultats par des hypothèses externes

Un premier élément d'explication de ces résultats est que dans les "réтины" spatialement variantes, la densité d'échantillonnage au centre est plus importante (d'un facteur 2 ou 4 selon le groupe) que la densité des réтины uniformes. Il n'est donc pas étonnant que les sujets soit plus précis dans le cas spatialement variant que dans le cas uniforme. Il est curieux néanmoins que dans ce cas on n'observe pas de différence significative de précision entre le groupe "Log2" et le groupe "Log4". Par ailleurs, cette hypothèse rend difficilement compte de la différence observée sur les temps de réponse. Nous ne voyons en effet aucun argument qui permettrait de faire l'hypothèse qu'un sujet bénéficiant d'une rétine uniforme de meilleure définition aille plus vite dans une tâche de pointé. Cette possibilité pourrait néanmoins être testée expérimentalement en utilisant des réтины uniformes de densités variables.

Une deuxième hypothèse concerne la question de la résolution de l'oreille, l'idée étant que puisque la densité des échantillons est plus forte au centre, l'oreille entend mieux les variations sonores produites par le dispositif. Cette deuxième hypothèse est fautive car la définition des sources sonores est faite avant la modification de la position des échantillons pour réaliser un échantillonnage spatialement variant. Les sources sonores sont donc aussi différenciables les unes des autres dans le groupe uniforme que dans les groupes spatialement variant. Ce qui est vrai en revanche, c'est que, vu l'effet de loupe associé à l'échantillonnage spatialement variant, deux points de l'image au centre de la rétine seront plus facilement discriminés dans ce dernier cas. Cette remarque est néanmoins totalement équivalente au point précédent.

La troisième hypothèse que nous proposons est que l'aspect spatialement variant de l'échantillonnage vient enrichir les contingences sensori-motrices liées au dispositif. Dans le cas uniforme en effet, lorsque je bouge la caméra, seule la fréquence et le panoramique du son change. Dans le cas spatialement variant en revanche, le nombre de sources activées change également, ainsi que la vitesse de variation du son, qui est maximale au centre de l'image. Cet enrichissement des contingences sensorimotrices est en particulier à même de fournir des indices supplémentaires concernant le centre du champ perceptif.

Si tel est le cas toutefois, la position des centres subjectifs des sujets dans les groupes spatialement variants doivent mieux coïncider avec le centre de l'image que dans le cas uniforme. Nous avons donc tracé sur un graphe la position des centres subjectifs des sujets en fonction de leur groupe ainsi que le centre objectif de l'image (fig 4.9). On observe qu'effectivement, la dispersion des centres subjectifs diminue entre les groupes "Unif", "Log2" et "Log4", et que ces derniers se rapprochent du centre objectif de l'image.

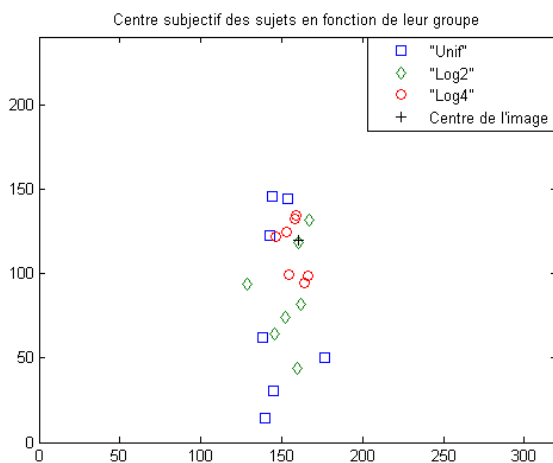


Figure 4.9 : Position du centre subjectif des sujets dans l'image, en fonction de leur groupe.

#### 4.1.4.4 Analyse des trajectoires de la cible

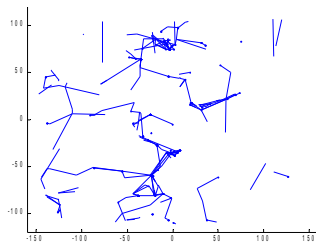
Afin de mieux cerner les différences de performances entre les sujets, nous nous sommes penchés sur la trajectoire de la cible entre son point de départ et son point d'arrivée. Afin de dégager d'éventuelles tendances, nous avons tracé l'ensemble des trajectoires de toutes les cibles pour chaque sujet. Dans l'idéal, ce graphe devrait ressembler à une étoile, centrée sur le centre subjectif du sujet.

Les graphes de tous les sujets sont fournis en annexe. Nous avons identifié dans ces graphes trois comportements typiques : le sujet "pressé", le sujet "prudent", et enfin le sujet "droit au but". La figure 4.10 présente deux exemples pour chaque comportement, mentionnant le sujet, son groupe, son temps de réponse, sa précision et le produit du temps de réponse par la précision, qui donne une estimation du compromis rapidité/précision.

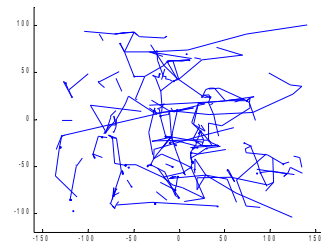
La comportement des sujets "pressés" (fig. 4.10, sujets 4 et 10) consiste à vrai dire à répondre le plus vite possible sans se soucier de la question de la précision. Leur temps de réponse défie donc toute concurrence, mais leur précision est proche du hasard.

Le comportement des sujets "prudents" (fig 4.10 sujets 5 et 13) consiste, pour déterminer le centre du champ perceptif, à se baser systématiquement sur les bords du champ de vision. Le sujet prudent fait sortir la cible d'un côté puis de l'autre afin de bien déterminer le centre du champ de la caméra. Ses mouvements sont donc essentiellement des mouvements horizontaux et verticaux, et il fait souvent plusieurs allé-retour avant d'arrêter son choix. Les performances de ce type de sujet sont relativement variables : le sujet 5 est assez lent et ne gagne pas en précision, le sujet 13 est, lui, assez rapide et d'une bonne précision.

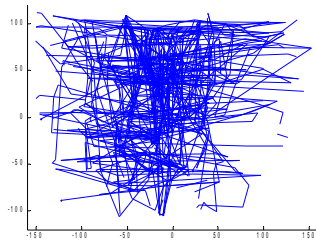
Le comportement des sujets "droit au but" (fig 4.10, sujets 6 et 8) produit des trajectoires qui se rapprochent de la trajectoire idéale en étoile. Ils vont directement placer la cible au centre de leur champ perceptif, et n'emploient pas nécessairement des chemins verticaux et horizontaux. Ce type de sujet est à la fois rapide et précis. Si l'on exclue les performances des sujets "pressés", qui finalement ne réussissent pas la tâche (leur précision correspond peu ou prou au hasard), ce sont ces sujets qui ont les meilleurs compromis rapidité/précision. Ils exploitent réellement le couplage sensori-moteur introduit par les variations de fréquence et de panoramique.



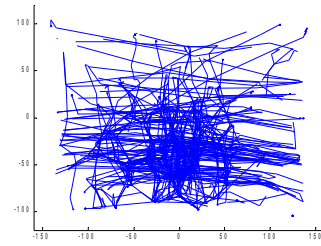
Sujet 4, Gr. "Unif", TR = 8, Pr = 54, TR\*Pr = 432



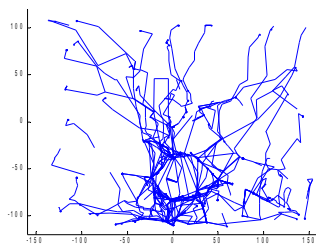
Sujet 10, Gr. "Log2", TR = 7, Pr = 56, TR\*Pr = 392



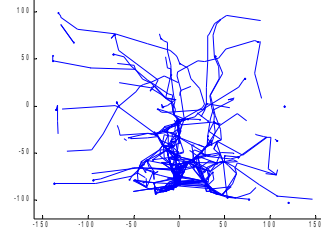
Sujet 5, Gr. "Unif", TR = 21, Pr = 46, TR\*Pr = 966



Sujet 13, Gr. "Log2", TR = 11, Pr = 10, TR\*Pr = 110



Sujet 6, Gr. "Unif", TR = 22, Pr = 32, TR\*Pr = 704



Sujet 8, Gr. "Log2", TR = 13, Pr = 8, TR\*Pr = 104

Figure 4.10 : position du centre subjectif des sujets dans l'image, en fonction de leur groupe. En haut, sujets "pressés", au milieu les sujets "prudents" et en bas les sujets "droit au but".

Bien sûr, nous nous sommes demandés si l'on pouvait dégager des tendances en fonction des groupes. En particulier, nous nous sommes demandés s'il n'y avait pas plus de sujets "prudent" dans le groupe "uniforme" que dans les groupes "spatialement variants". En effet, dans le cas du groupe "Unif", le centre du champ perceptif ouvert par le dispositif ne peut être défini qu'à partir de l'exploration des bords du champ, alors que dans le cas spatialement variant, d'autres indices y concourent. Il est néanmoins difficile de trancher sur cette question vu le faible nombre de sujets par groupe. Par ailleurs, les sujets ayant été prévenu de la méthode de codage, ils ont pu en déduire que le centre du champ perceptif correspondait à un son de fréquence donnée dont la position apparente est au centre de la tête, l'exploration dans ce cas n'est plus nécessaire, en tout cas au moins dans le sens horizontal (la notion de fréquence "centrale" n'a rien de naturel, contrairement à une source sonore d'azimut central).



#### 4.1.4.5 Les autres facteurs potentiels

Cette expérience est la première que nous avons réalisé avec le dispositif. De ce fait, le dispositif présentait certains défauts que nous avons corrigé par la suite. En particulier le fichier de "rétine" a été construit par la méthode des rectangles (voir chap. 3). Si cette méthode est meilleure que la technique d'origine consistant à choisir les centres des champs récepteurs aléatoirement dans l'image, certains défauts n'étaient pas totalement corrigés. En particulier, les rétines présentaient parfois des zones sous-échantillonnées où le sujet perdait momentanément la cible : le sujet 1 et le sujet 3 en ont fait la remarque à la fin de l'expérience.

Par ailleurs, l'application de la transformation logarithmique aux échantillons dans les groupes "Log2" et "Log4" produit des effets de bord qui ont tendance à rétrécir le champ de la caméra. Ce rétrécissement est de l'ordre de 10% pour les rétines "Log2" et de 20% pour les rétines "Log4". Il est évident que si le champ est plus petit, la tâche est plus facile. Néanmoins, cet élément peut difficilement rendre compte de la taille de l'effet observé : les performances varient de 50% entre le groupe uniforme et les groupes spatialement variant. Par ailleurs, on n'observe pas de différence significative de performance entre le groupe "Log2" et le groupe "Log4" alors qu'il y a la même différence en matière de rétrécissement du champ entre ces deux groupes qu'entre le groupe "Unif" et le groupe "Log2".

Enfin, nous n'avons pas songé pour cette première expérience à faire en sorte que l'opérateur ne sache pas dans quel groupe était le sujet. Le protocole est donc en simple aveugle. Toutefois, il est peu vraisemblable que l'opérateur ait influencé les résultats. Il ne pouvait en aucun cas influencer directement sur les résultats de l'expérience, puisque c'est le sujet qui déterminait le déroulement des essais.

Ces autres facteurs potentiels ne sont donc pas à même de rendre compte à eux seuls des résultats de l'expérience. Ils restent néanmoins à prendre en compte dans l'analyse.

#### 4.1.5 Conclusion

Dans cette expérience, nous avons testé, chez des sujets voyants équipés du dispositif TheVIBE, l'effet d'un échantillonnage spatialement variant biomimétique sur l'une des premières capacités accessibles lors du développement du système visuel humain : la capacité de réaliser des saccades vers une cible visuelle.

Pour cela, nous avons constitué trois groupes de sujets, dans lesquels l'échantillonnage était soit uniforme, soit spatialement variant avec un taux de magnification central de 2, soit spatialement variant avec un taux de magnification central de 4.

Nous avons placé les sujets devant un grand écran sur lequel nous affichons des cibles lumineuses sur un fond noir. Après une phase de prise en main d'environ 45 minutes, nous avons demandé aux sujets de "fixer la cible visuelle, c'est à dire de mettre la cible au centre du champ perceptif". Nous avons mesuré la durée mise par le sujet pour réaliser l'essai ainsi que la dispersion des positions de la cible à la fin de chaque essai.

Nous avons ainsi pu comparer les performances des sujets dans les trois groupes et constater que les sujets des groupes "spatialement variant" montrent des performances nettement supérieures à celles des sujets du groupe "uniforme". L'effet du taux de magnification dans les groupes "spatialement variant" est en revanche beaucoup plus discret.

Nous avons par la suite tenté d'analyser les raisons de ces différences de performance. Nous

pensons en particulier que l'existence d'un échantillonnage spatialement variant enrichit les contingences sensorimotrices du dispositif et que cela facilite la tâche de fixation. D'autres éléments sont toutefois à prendre en compte dans cette analyse et notamment l'effet de loupe associé à l'échantillonnage spatialement variant. Une analyse fine des trajectoires de la cible sur un plus grand nombre de sujets devrait permettre de mieux comprendre le phénomène observé.

## 4.2 Localisation-préhension à l'aide du dispositif TheVIBE

Forts des résultats obtenus dans la tâches de saccades, nous nous sommes demandés dans quelle mesure l'utilisation d'un échantillonnage spatialement variant pouvait influencer les performances des sujets dans une tâche plus complexe, en l'occurrence aller chercher une cible lumineuse dans une salle sombre. Cette tâche implique non seulement de pouvoir détecter la cible mais également de se mouvoir tout en gardant la cible en "contact" et enfin de la saisir.

Pour quantifier les performances du sujet, nous avons choisi de mesurer simplement le temps qu'il met pour trouver la cible. Il est évident que ce temps ne dépend pas seulement des performances du sujet mais aussi de la position initiale de la cible par rapport au sujet : si la cible est loin du sujet, il mettra plus de temps à s'en saisir que si elle est proche. Il s'agissait donc de construire un protocole permettant de contrôler ce paramètre, c'est à dire de faire en sorte que toutes les sessions soient équivalentes en "difficulté", de manière à pouvoir comparer les performances d'un sujet au fil des sessions, mais également à pouvoir comparer les sujets entre eux.

### 4.2.1 Méthode

Pour cela, nous avons mis en place un dispositif permettant de placer une cible lumineuse à différentes positions dans une salle. Le dispositif expérimental est constitué de trois tringles parallèles fixées au plafond, d'une corde munie de plusieurs noeuds, et d'une cible lumineuse : une balle translucide vide éclairée par une lampe torche (figures 4.11 et 4.12).



Figure 4.11 : un participant saisissant la cible



Figure 4.12 : le dispositif de fixation de la cible (au plafond)

Les tringles sont placées à environ 1 m 50 de distance séparant ainsi la salle en quatre bandes égales. La corde, munie d'un crochet à son extrémité, peut être placée à l'endroit souhaité sur l'une des tringles. On peut donc jouer sur la position latérale (choix de la tringle) et sur la distance de la

cible (position sur la tringle). La cible est ensuite fixée à la corde à l'un de ses noeuds. De cette manière, on peut jouer sur la hauteur de la cible.

#### 4.2.1.1 Appareillage

Les sujets portent un masque semi-rigide doublé de disques en coton pour garantir une occultation complète et efficace du champ visuel. Ils sont équipés du dispositif TheVIBE comprenant :

- Une webcam Quikcam Pro 5000 présentant un champ de 90° en diagonale fixée à l'arceau du casque juste au dessus de leur tête.
- Un casque stéréophonique Seinheiser HD 280 Pro.
- Un ordinateur de bureau Dell Latitude 620 équipé d'un processeur Intel Core 2 cadencé à 2 Ghz, et de 2 Go de mémoire vive.

L'image de la webcam est traitée par le logiciel TheVIBE qui génère un son transmis au sujet par les écouteurs.

##### **Version du logiciel :**

Pour cette expérience, le PC est équipé de la version 3.0 de TheVIBE.

Cette nouvelle version du logiciel, développée au laboratoire GIPSA en collaboration avec Gelu Ionescu, utilise les potentialités du codage vectoriel, ce qui élimine tout problème de fuite mémoire ou d'écrasement de variables. Cette nouvelle version a été développée dans le but d'y inclure un pré-traitement de l'image, comme nous le verrons au chapitre 5. Pour l'utilisation que l'on en fait ici, elle ne présente pas de différences majeures avec la version 2.0.

##### **Echantillonnage de l'image :**

Les fichiers de configuration bénéficient en revanche du savoir faire développé et présenté au chapitre 3.

En particulier, le fichier de configuration, unique pour chaque utilisateur, comprend 200 champs récepteurs dont les centres sont répartis grâce à un algorithme d'auto-organisation de type Kohonen (ch.3 §2.4). Chaque champ récepteur est composé de 10 échantillons distribués aléatoirement selon une distribution de probabilité gaussienne, centrée sur le centre du champ récepteur et d'écart-type 5 pixels.

Pour créer des rétines spatialement variantes, on applique une transformation logarithmique aux échantillons d'une rétine "uniforme". Compte tenu des résultats précédents, le taux de magnification dans le groupe "Log" a été fixé à 3.

L'utilisation de la méthode de Kohonen a pour avantage :

1. De réduire considérablement les zones sous-échantillonnées par le dispositif. En particulier, aucun sujet n'a mentionné de "trous" dans son champ perceptif.
2. D'échantillonner une zone circulaire, ce qui permet de s'affranchir des effets de bords que l'on obtenait dans l'expérience précédente en appliquant une transformation logarithmique aux échantillons. De la sorte, la zone échantillonnée dans les rétines "spatialement variantes" est exactement la même que celle dans les rétines "uniformes" (fig 4.13).

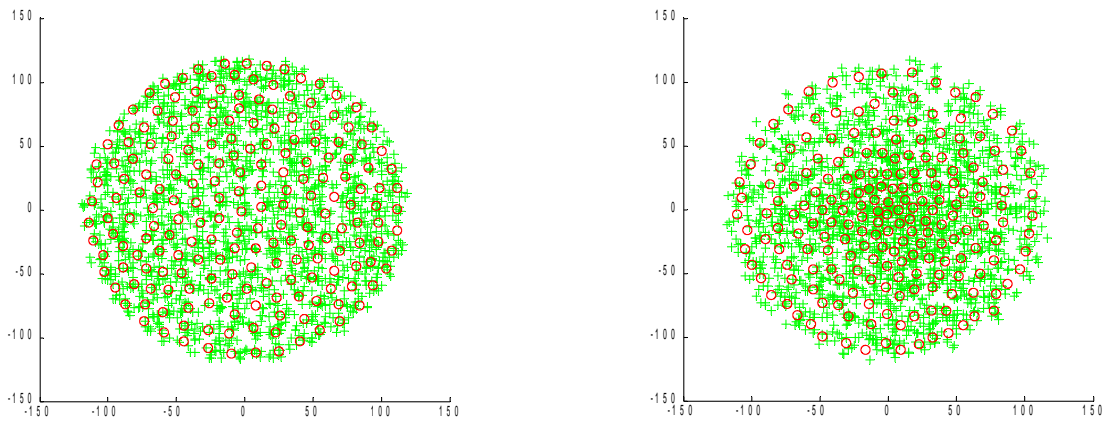


Figure 4.13 : répartition des champs récepteurs et des échantillons dans l'expérience. A gauche, échantillonnage uniforme, à droite, échantillonnage spatialement variant.

Par ailleurs, la distribution des échantillons dans l'image est corrigée pour tenir compte de l'optique de la caméra (ch. 3 §2.6). Cette déformation est toutefois faible dans notre cas.

### **Mapping visuo-auditif :**

Tout comme dans l'expérience précédente, le mapping visuo-auditif est de type "Euclidien" (ch. 3 §3.5.1) pour tous les sujets. Les échelles utilisées en revanche sont différentes et mieux adaptées à la physiologie de l'oreille.

- Le codage de l'activité des champs récepteurs utilise une échelle d'amplitude logarithmique (voir Ch. 3 §3.3.1)
- Le codage du panoramique utilise une différence d'intensité sonore entre les deux oreilles basée également sur une échelle logarithmique (voir Ch. 3 §3.3.2)
- Le codage de la tonie utilise une échelle en Bark, dans l'intervalle 300-3000 Hz (voir Ch. 3 §3.3.3) soit environ 3-13 Barks.

Rappelons encore une fois que l'attribution des paramètres sonores des sources associées aux champs récepteurs est faite à partir de la position des champs récepteurs avant une éventuelle modification de la position des échantillons pour réaliser un échantillonnage spatialement variant.

### **Enregistrement de l'activité des champs récepteurs :**

L'activité des champs récepteurs est enregistrée dans un fichier extérieur une fois toutes les deux images, soit environ 12 fois par secondes.

### **Environnement :**

Les sujets sont placés sur un siège à l'extrémité d'une pièce vide et sombre d'environ 6 mètres par 9 face au dispositif de fixation de la cible (fig. 4.14).

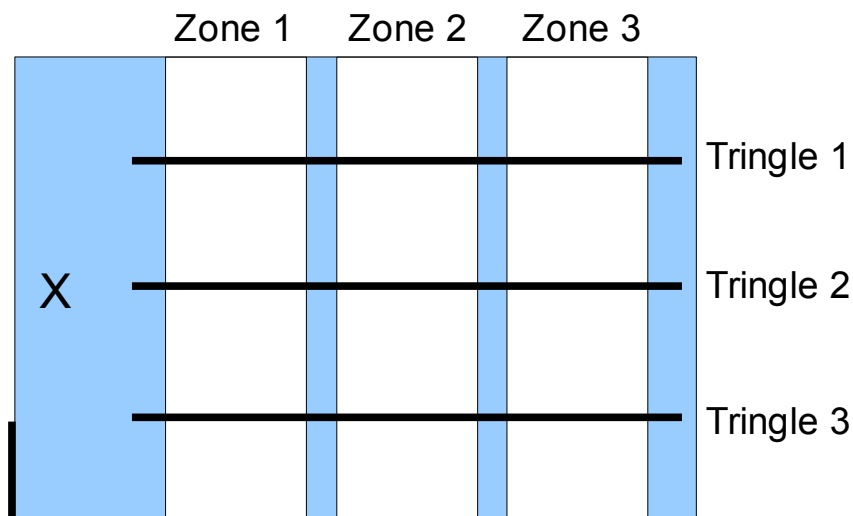


Figure 4.14 : disposition de la salle. Le sujet (croix) est assis face au dispositif de fixation. La position de la cible est définie par un numéro de zone et un numéro de tringle. Un troisième numéro correspond à la hauteur de la cible (1 : en haut, 2 : au milieu, 3 : en bas)

## 4.2.2 Protocole

Le protocole de cette expérience est particulièrement simple : il est composé de trois sessions de neuf essais chacune, séparées par une pause de 5 minutes environ.

Au début de l'expérience, le sujet est placé sur le siège et équipé du dispositif. Il est informé du fait que le dispositif traduit les images de la caméra en son. Aucune information n'est fournie quant au processus de codage.

A chaque essai, l'opérateur place une cible dans la salle et vient se placer à gauche du sujet. Le sujet doit alors tout d'abord repérer la cible, c'est à dire la mettre dans son champ de vision, sans bouger de son siège. Lorsque c'est le cas, le sujet doit se lever et aller s'emparer de la cible. On mesure le temps mis par le sujet pour s'emparer effectivement de la cible, à partir du moment où il se lève de son siège. A la fin de l'essai, l'opérateur ramène le sujet à son siège.

Le point important du protocole réside en fait dans le choix des positions de la cible lors des trois sessions. En effet, la durée d'un essai dépend non seulement des performances du sujet mais aussi de la position de la cible par rapport à la position initiale du sujet. Il est donc nécessaire de contrebalancer cet effet. Par ailleurs, le sujet peut apprendre la position d'une cible. Afin d'éviter cela, nous avons fait en sorte qu'aucune cible ne soit présentée deux fois. Pour cela nous avons considéré trois possibilités pour la position latérale de la cible (tringle 1, 2 ou 3), trois positions en profondeur (zone 1, 2, ou 3), et enfin trois hauteurs de la cible (Haut, Milieu, Bas). La figure 4.14 illustre ce propos. Le tout donne bien évidemment  $3^3 = 27$  positions possibles que nous avons répartis sur les trois sessions.

Dans la mesure où nous souhaitons comparer les performances des sujets dans chaque session, un contrebalancement aléatoire était trop risqué, en particulier parce qu'un décalage dans une session par exemple une majorité de cibles proches, aurait induit un décalage dans les autres sessions : une majorité de cibles lointaines dans le cas présent. Pour équilibrer la difficulté des sessions, nous faisons en sorte que toutes les hauteurs de cibles soient présentées dans chaque zone et pour chaque tringle. Cela revient à dire qu'aucune cible dans une zone donnée ou sur une tringle donnée, n'a la même hauteur que ses voisines. En pratique, on commence par la première session

en procédant zone par zone et tringle par tringle en choisissant à chaque fois une hauteur de cible, sans possibilité de remise. On choisit d'abord une hauteur pour la cible placée sur la tringle 1 dans la zone 1. Parmi les deux restantes, on choisit la hauteur de celle de la tringle 2. Celle de la tringle 3 est alors déterminée. De la sorte, il y a bien une cible basse, une cible moyenne et une cible haute dans la zone 1. On passe ensuite à la zone 2. Cette fois le choix est plus limité car, pour qu'aucune cible sur la même tringle n'ait la même hauteur, il ne reste que deux configurations possibles. La configuration de la zone 3 est ensuite entièrement déterminée. Pour les deux autres sessions, on permute soit deux indices de tringle, soit deux indices de zone. De la sorte, l'équilibre est toujours respecté, et aucune cible n'est présentée deux fois au cours de l'expérience. Une fois les positions des cibles déterminées dans chaque session, leur ordre par session est bien sûr aléatoire.

### 4.2.3 Participants et plan d'expérience

Seize sujets, 10 hommes et 6 femmes, voyants et sans problèmes auditifs avérés ont pris part à l'expérience. Leur âge s'étend de 21 à 65 ans, avec une moyenne de 30 ans et un écart-type de 12 ans. Aucun d'entre eux n'avait utilisé ce type d'appareil auparavant. Les sujets ont été répartis dans deux groupes (Unif, Log) appariés pour leur sexe et pour leur âge (test de Student,  $p = 0.71$ ).

Le plan expérimental est entièrement inter :  $S_{16} < E_2 > * A_3$

où le facteur indépendant E correspond au type d'échantillonnage de l'image dans le dispositif : uniforme ("Unif") ou spatialement variant ("Log") et le facteur indépendant A correspond à la session d'apprentissage.

### 4.2.4 Résultats

Le tableau 3 présente le temps total des sujets, en secondes, pour les trois sessions de l'expérience (respectivement A1, A2 et A3), ainsi que leur groupe.

La figure 4.15 présente les moyennes des sujets en fonction de la session, pour le groupe "Unif" et pour le groupe "Log".

Pour déterminer la méthode statistique appropriée à l'analyse de ces données, nous avons tout d'abord testé l'homogénéité des variances de tous les groupes à l'aide d'un test de Bartlett (1937). Les variances des différents groupes ne sont pas significativement différentes les uns des autres (Bartlett test,  $p = .07$ ). Nous avons ensuite testé la normalité des distributions en fonction des groupes. Seule la distribution des valeurs du groupe "Unif" dans la session 3 est significativement différente d'une distribution gaussienne ( $p < .001$ ).

Nous avons donc cherché l'existence d'un éventuel outlier dans cette série à l'aide d'un test de Grubbs (1969). Ce dernier signale effectivement un outlier : le sujet 16 dont le temps moyen est de 442 secondes (test de Grubbs,  $p < 0.001$ ).

Ce sujet a par ailleurs une courbe très différente de tous les autres sujets puisqu'il est le seul à voir son temps total augmenter avec la session. Ce sujet avait par ailleurs mentionné, lors du test, des problèmes de coordination motrices à la naissance. Nous l'avons donc écarté de l'analyse statistique.

Sans ce sujet, aucune distribution ne présente de résultat significatif au test de Shapiro-Wilk.

Pour tester l'effet de l'apprentissage, nous avons comme précédemment effectué deux contrastes, l'un comparant la première session aux deux suivantes (C1), et l'autre comparant les deux dernières sessions entre elles (C2). Les deux contrastes sont nettement significatifs (test de Student,  $p < 0.001$ ).

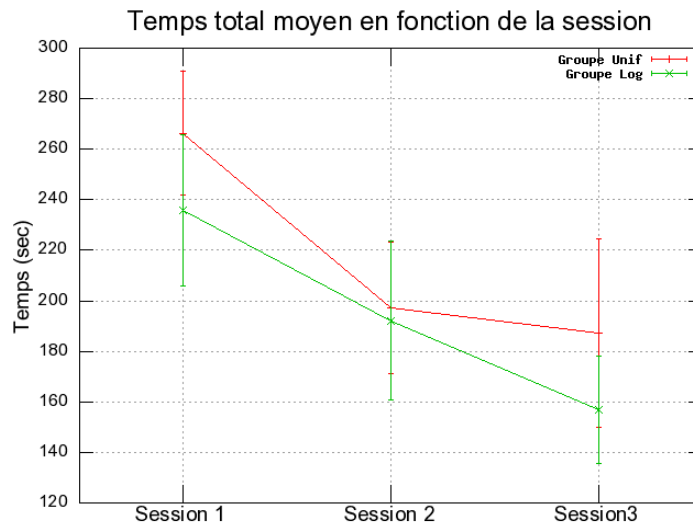


Figure 4.15 : Temps totale moyen des sujets par groupe en fonction de la session

pour C1, et  $p < 0.01$  pour C2).

Pour tester l'effet de l'échantillonnage spatialement variant sur les performances globales des sujets, nous avons comparé les temps totaux des sujets sur les trois sessions de l'expérience entre le groupe "Unif" et le groupe "Log". Les temps totaux du groupe "Log" sont en revanche non significativement différents de ceux du groupe "Unif" (test de Student,  $p = 0.91$ ).

## 4.2.5 Discussion

Dans cette expérience, nous avons testé la capacité de sujets équipés du dispositif TheVIBE à aller chercher une cible lumineuse fixée à différents endroits dans une salle sombre. Nous avons également comparé les performances des sujets utilisant un dispositif présentant un échantillonnage spatialement variant de l'image et des sujets pour lesquels l'échantillonnage de l'image était uniforme.

### 4.2.5.1 Principaux résultats

Le premier constat de cette expérience est que les sujets sont bien capables d'aller chercher la cible dans la salle. Le temps moyen pour aller chercher cette cible, après deux sessions d'apprentissages, est de l'ordre de 20 secondes, ce qui est bien sûr plus élevé que chez un sujet voyant dans ces conditions, mais reste néanmoins acceptable pour une utilisation en pratique.

Par ailleurs, les performances des sujets s'améliorent significativement entre la première session et les deux suivantes. Cet effet n'est pas étonnant mais il nous renseigne sur la capacité du protocole expérimental à mettre des différences de performance en évidence.

Enfin, nous ne trouvons pas de différence de performances entre le groupe "Log" et le groupe "Unif". Ce dernier point est lui particulièrement étonnant au regard des résultats obtenus dans l'expérience précédente sur la capacité à fixer une cible avec le dispositif. Une raison possible est que la taille de la zone d'échantillonnage est beaucoup plus petite dans cette expérience que dans la précédente. En effet, le dispositif échantillonnait dans cette expérience une fenêtre circulaire de 240 pixels de diamètre (fig 4.13), ce qui réduit le champ à  $58^\circ$  en diagonale, contre  $90^\circ$  précédemment.

Dans ces conditions, l'effet d'un échantillonnage spatialement variant est peut-être minime.

Une autre raison potentielle est que nous avons changé d'ordinateur entre les deux expériences, celui dans cette expérience étant plus performant que le précédent. De ce fait, le décalage temporel entre l'image et le son correspondant (le "lag") est sans doute réduit (bien que nous n'ayons pas de moyen de le mesurer en pratique). Il est tout à fait possible que l'effet d'un échantillonnage spatialement-variant soit en interaction avec ce décalage : plus le décalage est grand et plus la nature des contingences sensori-motrices influe sur les performances du sujet. Si le décalage est très faible, le sujet peut rapidement compenser une différence dans les contingences sensorimotrices par des mouvements, notamment des mouvements de balayages.

Dans ce cas, nous devrions observer une augmentation de la trajectoire totale de la cible dans le champ de la caméra. Pour tester cette hypothèse, nous avons calculé, pour chaque sujet, la longueur totale de la trajectoire de la cible dans le champ de la caméra lors de la dernière session d'apprentissage. Les résultats sont présentés dans le tableau 4.

Tableau 4 : Longueur totale du parcours lors de la dernière session d'apprentissage.

Sujet	Groupe	Parcours Tot A3	Sujet	Groupe	Parcours Tot A3
9	Log	323070	2	Unif	616010
10	Log	376490	5	Unif	214900
11	Log	360290	6	Unif	317160
12	Log	235600	7	Unif	298910
13	Log	350630	8	Unif	310790
15	Log	169380	1	Unif	433140
17	Log	464180	14	Unif	481310
18	Log	281070	16	Unif	465570
	<b>Moyenne</b>	<b>320088,75</b>		<b>Moyenne</b>	<b>392223,75</b>
	<b>Erreur standard</b>	<b>32147,85</b>		<b>Erreur standard</b>	<b>45803,86</b>

Après avoir testé l'homogénéité des variances (Test de Bartlett,  $p = 0.33$ ), et la normalité des mesures (Test de Shapiro-Wilk,  $p = 0.97$  pour le groupe "Log", et  $p = 0.58$  pour le groupe "Unif"), nous avons testé la significativité de la différence des moyennes entre les groupes. Cette différence n'est pas significative (Test de Student,  $p = 0.22$ ). Il n'est donc pas possible de conclure quant à cette hypothèse. Celle-ci reste néanmoins plausible.

#### 4.2.5.2 Analyse du comportement des sujets

Dans un deuxième temps, nous nous sommes intéressés au comportement des sujets pour tenter de mettre en évidence des modifications liées à l'apprentissage de la tâche et du dispositif.

Pour mettre en évidence des tendances générales, nous avons comme précédemment tracé le parcours de la cible pour chaque sujet et pour chaque session, incluant donc les données sur neuf essais. Le type de graphique obtenu est présenté figure 4.16.

Ce type de graphique, utile pour l'expérience précédente, est cette fois-ci peu lisible du fait des mouvements très nombreux du sujet. Nous avons donc opté pour deux autres représentations. La première (fig. 4.17) présente la densité des positions de la cible dans le champ de la caméra pour une session d'apprentissage donnée. Ceci permet de voir en particulier si le sujet arrive à maintenir une fixation sur la cible, donnant dans ce cas une zone à forte densité. L'évolution de ce graphique renseigne donc sur les capacités du sujet à maintenir le contact avec la cible, étape importante de l'immersion dans le dispositif.



La deuxième (fig. 4.18) est un histogramme des orientations des mouvement de la cible dans l'image calculés entre deux mesures consécutives de la position de la cible dans le champ de la caméra. Encore une fois, les données considérées concernent une session complète. L'angle de  $0^\circ$  correspond à un mouvement de balayage horizontal, et celui de  $90^\circ$  à un mouvement de balayage vertical (haut-bas) de l'image. Le sens de balayage n'est pas considéré. Ce graphe renseigne donc sur les moyens employés par le sujet pour maintenir la cible en contact.

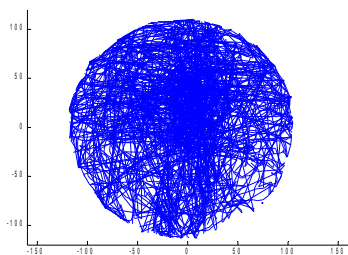


Figure 4.16 : trajectoire de la cible durant une session d'apprentissage.

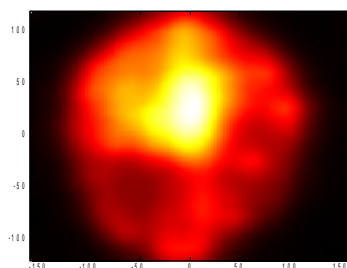


Figure 4.17 : densité des positions de la cible durant une session d'apprentissage

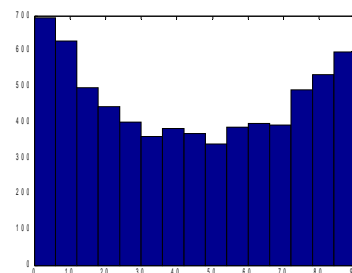


Figure 4.18 : histogramme des orientations des mouvements de la cible dans l'image

Les graphiques de densité et les histogrammes par session pour chaque sujet sont fournis en annexe. S'il est difficile de discerner une tendance générale chez tous les sujets, certains sujets présentent une évolution assez caractéristique. Le sujet 10 du groupe "Log" par exemple, qui présente l'une des courbes d'apprentissage les plus marquées. La figure 4.19 présente l'évolution de la densité des points et de l'histogramme des orientations au cours des trois sessions pour ce sujet.

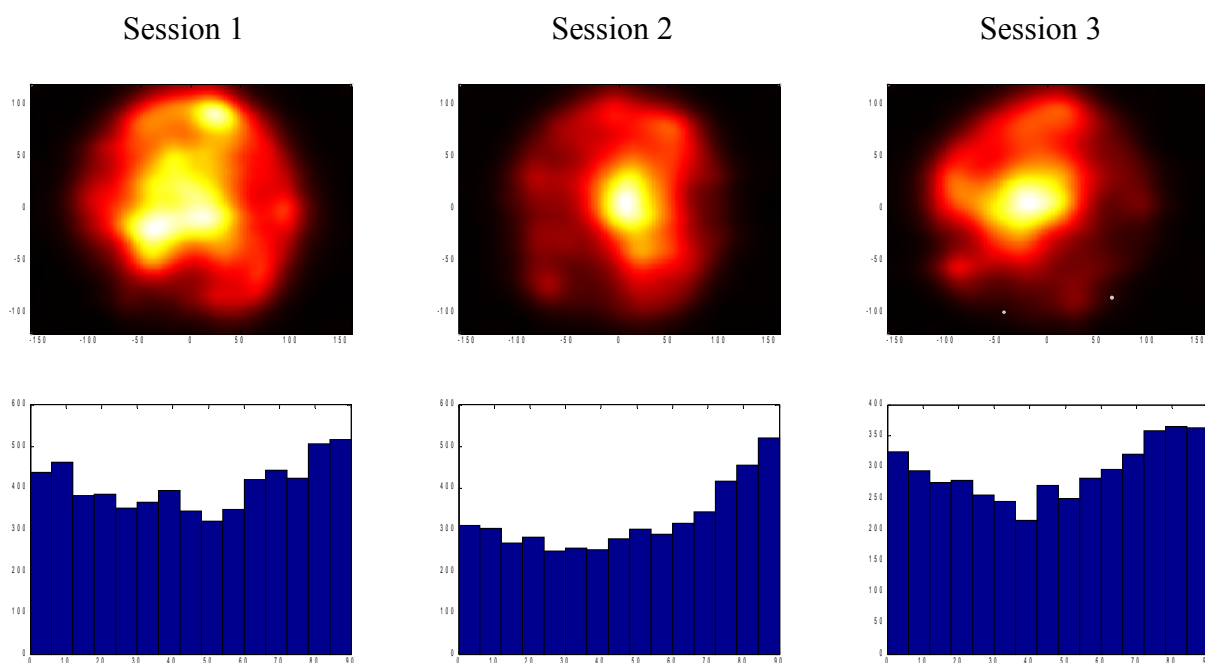


Figure 4.19 : évolution du comportement du sujet 10 (groupe "Log") lors des trois sessions d'apprentissage.

Au niveau de la densité des positions de la cible dans l'image, on note chez ce sujet une réduction

de la dispersion des points qui se concentrent vers le centre. Le sujet semble donc "fovéaliser" de mieux en mieux la cible. Par ailleurs, au fil des sessions, les mouvements de balayage semblent se concentrer sur les directions horizontales et verticales. Ce dernier aspect est plus délicat à interpréter, en particulier parce que ce type de comportement fait écho au comportement des sujets "prudents" de l'expérience précédente, qui sont moins efficaces en général que les sujets "droit au but", qui n'ont pas d'orientation de mouvement privilégiée. Toutefois, cette préférence vers les orientations horizontales et verticales reste relative car les autres orientations sont également bien représentées. Ce type de comportement se retrouve chez d'autres sujets du groupe "Log" : les sujets 9, 11, et 18.

Dans le groupe "Unif", ce type de comportement est plus difficile à discerner. Les sujets 2 et 7 semblent s'en rapprocher mais, en particulier, la distribution finale n'est pas toujours unipolaire, comme le montre l'exemple du sujet 7. On remarque en particulier sur le sujet 7 la dispersion finale au niveau de l'axe vertical, qui signale une difficulté particulière à maintenir la cible en contact à l'aide de la fréquence du signal (fig. 4.20).

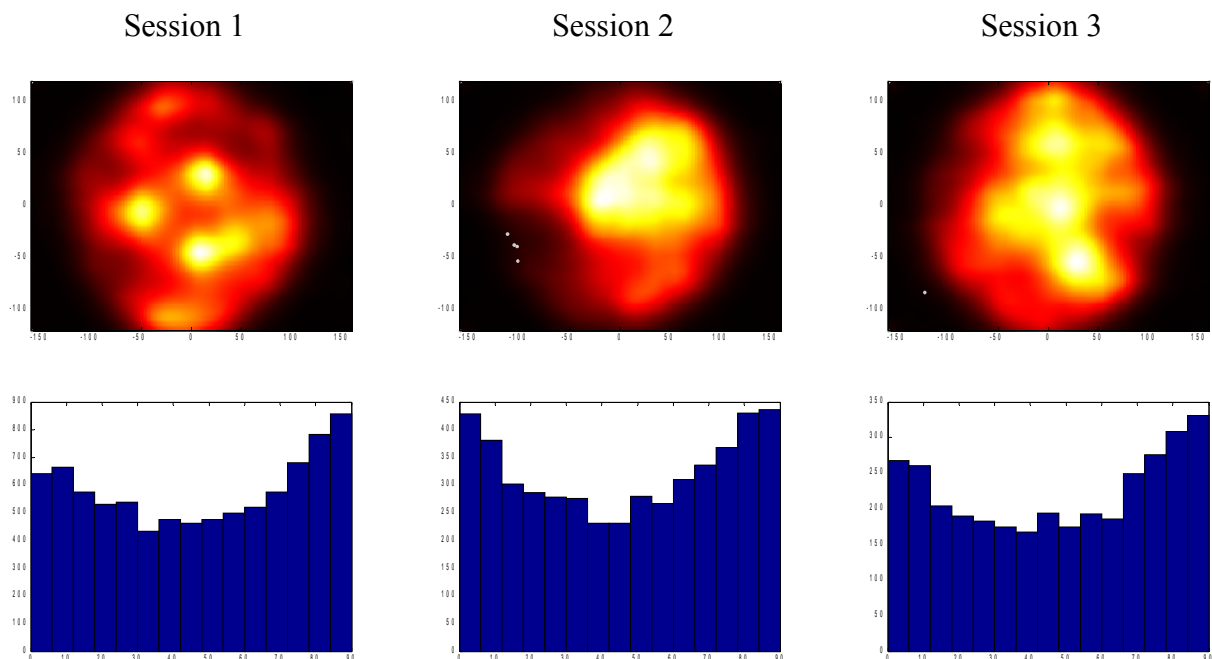


Figure 4.20 : évolution du comportement du sujet 7 (groupe "Unif") lors des trois sessions d'apprentissage.

On notera enfin le comportement très particulier de l'un des sujets du groupe "Log" : le sujet 15, présenté figure 4.21. La densité des cibles chez ce sujet est très centrale et très peu dispersée dès la première session (les points blancs sont des points de mesure considérés comme outliers tant la distribution est serrée). Par ailleurs, l'histogramme des orientations de mouvements est plat : il ne présente pas de direction privilégiée.

Ces différences marquées avec le reste des sujets correspondent en fait à une différence de stratégie. A l'issue de l'expérience, nous lui avons demandé de décrire sa stratégie et si possible de proposer une hypothèse quand au codage de l'information. Ce sujet avait répondu qu'il ne pouvait pas formuler d'hypothèses quant au codage, mais qu'il se basait sur le volume sonore pour garder la cible en contact. En effet, si l'échantillonnage dans le dispositif est spatialement variant, le volume sonore varie en fonction de la position de la cible dans l'image, car elle active plus ou moins de

champs récepteurs. Ce dernier sujet s'est donc basé uniquement sur cette information. Remarquons toutefois que cette stratégie n'est pas particulièrement efficace, puisque le sujet mettait en moyenne 30 secondes pour aller chercher une cible, contre 20 secondes en moyenne pour le groupe.

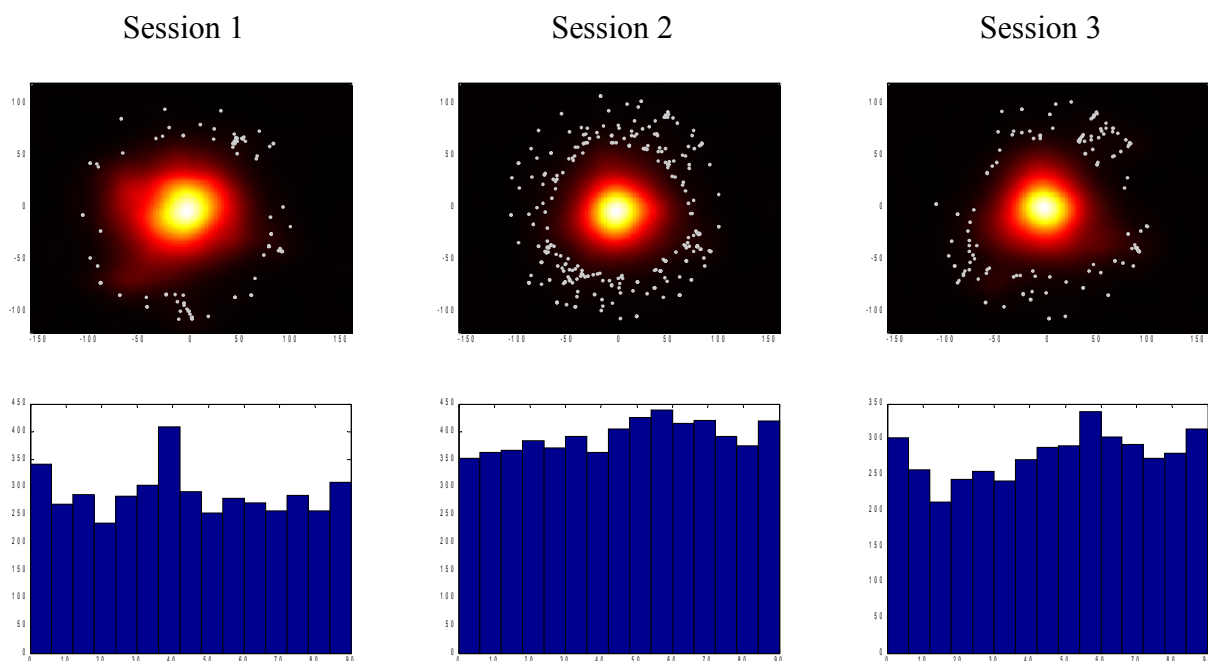


Figure 4.21 : évolution du comportement du sujet 15 (groupe "Log") lors des trois sessions d'apprentissage.

## 4.2.6 Conclusion

Dans cette expérience, nous avons étudié l'effet de l'apprentissage et d'un échantillonnage spatialement variant de l'image dans le dispositif de substitution sur la capacité des sujets à aller chercher une cible lumineuse dans une pièce sombre.

Nous avons montré en particulier que les performances s'améliorent avec la pratique, mais nous n'avons pas pu mettre en évidence un effet de l'échantillonnage. Il se peut toutefois que cet effet soit compensé par le sujet, en augmentant le nombre de mouvements avec la caméra, mais cela reste à confirmer.

Par ailleurs une analyse des trajectoires de la cible dans le champ de la caméra permet de proposer des hypothèses quant au processus d'apprentissage avec le dispositif. En particulier, la modification des cartes de densité de position de la cible permet d'illustrer la phase de "contact", première étape dans l'immersion d'un dispositif, dans laquelle le sujet apprend les "régularités sensorimotrices nécessaires à stabiliser et à maintenir le contact perceptif avec le stimulus".

## 4.3 Conclusion générale

Dans ce chapitre nous nous sommes penchés sur le cas très simple de la perception d'une cible lumineuse dans un environnement sombre avec le dispositif TheVIBE. Cela nous a permis de mettre en évidence un apprentissage des sujets, aussi bien dans une tâche de fovéalisation de la cible que dans une tâche consistant à aller chercher la cible dans la pièce.

Nous avons par ailleurs testé les modifications induites par un échantillonnage spatialement variant de l'image de la caméra. Dans le cas de la tâche de fovéalisaton, nous avons effectivement mis en évidence un effet de cet échantillonnage, qui améliore la précision et la rapidité des sujets en amoindrissant le temps d'apprentissage. Nous proposons que cet effet soit lié à un enrichissement des contingences sensori-motrices liées à ce type d'échantillonnage.

Parallèlement, nous avons étudié les stratégies employées par les sujets et identifié plusieurs types de comportement. Aucun lien entre ces stratégies et l'échantillonnage spatialement variant n'a toutefois pu être clairement identifiée.

Pour la tâche consistant à aller chercher la cible dans la salle, nous n'avons pas pu mettre un tel effet en évidence. En étudiant la trajectoire de la cible dans le champ de la caméra, nous avons toutefois identifié des différences dans la capacité à maintenir la cible en fovéa, sans toutefois pouvoir établir clairement une corrélation avec le type d'échantillonnage dans le dispositif.

Si ces expériences nous ont permis de mettre en évidence la capacité des sujets à apprendre l'utilisation du dispositif, ainsi que le potentiel de ce type de dispositif pour étudier la question de la genèse perceptive, la situation choisie reste néanmoins très particulière : il n'y a qu'un seul objet lumineux dans l'environnement. On est donc très loin des conditions dans lequel notre système visuel travail usuellement. Pour cette raison nous avons choisi de démarrer une nouvelle étude qui, à l'inverse de celle-ci, est réalisée dans un environnement complexe et non contrôlé : un parcours en extérieur, dans lequel nous avons testé l'impact du dispositif sur la mobilité des sujets. Cette étude fait l'objet du prochain chapitre.

## Références

Auvray, M. (2004), 'Immersion et perception spatiale : L'exemple des dispositifs de substitution sensorielle', PhD thesis, Ehess, Paris.

Auvray, M.; Hanneton, S.; Lenay, C. & O'Regan, K. (2005), 'There is something out there: distal attribution in sensory substitution, twenty years later', *Journal of Integrative Neuroscience* **4**(4), 505.

Bartlett, M. S. (1937), 'Properties of sufficiency and statistical tests', *Proceedings of the Royal Statistical Society Series A* **160**, 268-282.

Brainard, D. (1997), 'The psychophysics toolbox', *Spatial Vision* **10**(4), 433-436.

Epstein, W.; Hughes, B.; Schneider, S. & Bach-y-Rita, P. (1986), 'Is there anything out there? A study of distal attribution in response to vibrotactile stimulation.', *Perception* **15**(3), 275-284.

FNRO (2009), 'Évaluation clinique des étapes de la maturation visuelle chez l'enfant', <http://www.documents.strabisme.net/page1151/page1154/page1154.html>.

Grubbs, F. (1969), 'Procedures for detecting outlying observations in samples', *Technometrics*, 1-21.

Mazeau, M.(Paris), M., ed. (1995), *Déficits visuo-spatiaux et dyspraxies de l'enfant atteint de lésions cérébrales précoces: du trouble à la rééducation*.

Shapiro, S. & Wilk, M. (1965), 'An analysis of variance test for normality (complete samples)', *Biometrika* **52**(3,4), 591-611.

Wilcoxon, F. (1945), 'Individual comparisons by ranking methods', *Biometrics Bulletin*, 80--83.

Zanlonghi (2001), 'La vision de l'enfant',  
[http://www.ophtalmo.net/bv/GP/IndexGP/G/Vision\\_normale/vn-enf1.htm](http://www.ophtalmo.net/bv/GP/IndexGP/G/Vision_normale/vn-enf1.htm).

## Chapitre 5 : Evaluation de l'efficacité du dispositif de substitution pour la mobilité

### 5.1 Pourquoi la mobilité?

Dans ce chapitre nous allons nous intéresser plus particulièrement à la question de l'intérêt potentiel de la substitution sensorielle pour l'assistance à la mobilité des sujets, et ceci pour plusieurs raisons:

1 – La résolution d'un dispositif de substitution est très faible comparée à la résolution de l'œil humain. De ce fait, nous pensons que la substitution est peu appropriée aux tâches requérant une bonne acuité visuelle. Par ailleurs, d'après une étude de West *et al.* (2002) chez les sujets malvoyants, la mobilité n'est affectée que très tard par une diminution de la sensibilité visuelle. Il signale que certains sujets ayant une acuité inférieure à 1 logMAR (équivalent à 1 dixième sur l'échelle usuelle en France, c'est à dire à ne pas reconnaître les plus grosses lettres du tableau de Monoyer à 5 m) ne présentent pas de handicap dans ce domaine. Il est donc vraisemblable que, malgré une résolution très faible, le dispositif de substitution puisse être utile à un non-voyant dans une tâche de mobilité.

On trouve également dans la littérature une série d'études visant à quantifier la résolution minimale pour qu'une prothèse visuelle soit utile dans une tâche de mobilité (Cha *et al.*, 1992 ; Dowling *et al.*, 2005 ; Dagnelie *et al.*, 2007). Les auteurs utilisent pour cela un casque de réalité virtuelle dans lequel ils projettent l'image issue de la caméra. Cette image est pixelisée et filtrée de manière à simuler l'image obtenue au moyen d'une prothèse visuelle. Ils étudient ensuite la mobilité de sujets voyants équipés de ce dispositif. Selon l'étude de Dagnélie, une prothèse de 60 points présente déjà un intérêt pour la mobilité. L'étude de Cha utilise un dispositif de 625 points. Quant à Dowling, son dispositif est constitué de 2500 points.

2 – L'incapacité à se déplacer de manière autonome est l'un des aspects les plus handicapant dans la vie quotidienne d'un non-voyant. C'est en tout cas la conclusion d'une étude réalisée par le Ministère de la Santé Français (Sander 2005) : "La réduction de l'autonomie dans la vie quotidienne est l'une des principales conséquences des déficiences visuelles et plus d'un déficient visuel sur deux (56%) déclare une incapacité sévère concernant la mobilité et les déplacements". La conception d'un dispositif aidant à la mobilité des non-voyants présente donc un enjeu social évident.

3 – La mobilité est une tâche particulièrement intéressante, à la fois du point de vue des théories actives de la perception et du point de vue du traitement de l'information dans le système visuel. La mobilité est considérée par Gibson comme l'une des principales affordances d'un environnement (Gibson, 1986). La notion de "flux optique" est d'ailleurs particulièrement pertinente dans cette tâche, puisque ce flux est à même de renseigner le sujet sur la direction et la vitesse de son déplacement. En ce sens, les contingences sensori-motrices mises en jeu dans une tâche de mobilité sont beaucoup plus riches que dans une tâche comme la localisation d'une cible.

La mobilité est par ailleurs intéressante du point de vue du traitement de l'information dans le système visuel. Nous avons vu au chapitre 2 qu'on avait identifié au moins deux voies dans le système visuel primaire, la voie Parvocellulaire et la voie Magnocellulaire, dont les propriétés diffèrent en terme de comportement spatio-temporel. Ces voies ont vraisemblablement un rôle fonctionnel différent, la voie Parvo étant plutôt dédiée aux tâches de reconnaissance des objets, de

lecture, etc... et la voie Magno plutôt dédiée au mouvement. On peut donc légitimement supposer que la voie Magno a des propriétés intéressantes vis à vis de la mobilité. De là l'idée d'intégrer dans notre dispositif une simulation du traitement de l'information de la voie Magnocellulaire, et de tester son impact sur l'efficacité du dispositif.

## 5.2 Evaluer la mobilité

La mobilité est définie par Foulke (Shringledecker et Foulke, 1978) comme la capacité à se déplacer d'un endroit à un autre de manière gracieuse, sûre, confortable, et autonome<sup>8</sup>. Cette définition très générale a par la suite été complétée par Jansson (1986) qui identifie quatre composantes essentielles de la mobilité<sup>9</sup> :

- 1 – Garder l'équilibre
- 2 – Marcher vers un but
- 3 – Marcher suivant une ligne directrice
- 4 – Eviter les obstacles

Le point 2, marcher vers un but, fait référence aux capacités d'orientation du sujet, qui implique qu'il soit capable de repérer sa propre position et celle de son but. La vision est bien sûr très impliquée dans la capacité d'orientation. Il peut s'agir de repérer des "landmarks", mais la plupart du temps, sur des trajets connus, c'est l'intégralité de la scène qui est utilisée pour s'orienter (bâtiments, allées, etc...) et ce de manière parfaitement automatique.

Le point 3, marcher suivant une ligne directrice, fait notamment référence au problème de "veering" observé chez les sujets mal et non-voyants. Sur un parcours rectiligne, ces derniers ont tendance à dévier de leur trajectoire. Cette déviation est de l'ordre de 1m50 pour 10 m (Kallie *et al.*, 2007), ce qui pose de nombreux problèmes en environnement urbain en particulier lorsque le sujet s'éloigne des bâtiments, qui lui servent souvent de repères.

Le point 4, éviter les obstacles, considère à la fois les obstacles statiques du parcours et les obstacles en mouvement (autres piétons, voitures, etc...)

Compte tenu de la résolution du dispositif TheVIBE, qui ne permet pas *a priori* de distinguer des objets de petite taille, nous nous sommes focalisés sur les trois premières composantes de la mobilité. En particulier, nous n'avons pas considéré la question de détecter des obstacles sur le parcours, ou de reconnaître des landmarks de petite taille (panneaux, flèches, etc..).

### 5.2.1 Evaluer un dispositif de substitution : la question de l'apprentissage

La méthode la plus courante pour évaluer les capacités de mobilité d'un sujet consiste à construire un parcours que ce dernier devra effectuer (pour une revue des méthodes d'évaluation de la mobilité, voir Leat et Lovie-Kitchin, 2006). On mesure alors le temps de parcours ainsi que le nombre "d'erreurs de mobilité" commises par le sujet, comme par exemple une pause, ou un contact avec un obstacle. Ces erreurs sont en général comptabilisées par un observateur spécialisé dans le domaine de l'orientation et de la mobilité (O&M).

L'évaluation d'un dispositif de substitution pose toutefois une question supplémentaire : celle de

<sup>8</sup>"ability to travel between locations gracefully, safely, comfortably and independently"

<sup>9</sup> "1- Keeping balance ; 2 - Walking toward a goal ; 3 - Walking along a guideline ; 4 - Walking around obstacles"

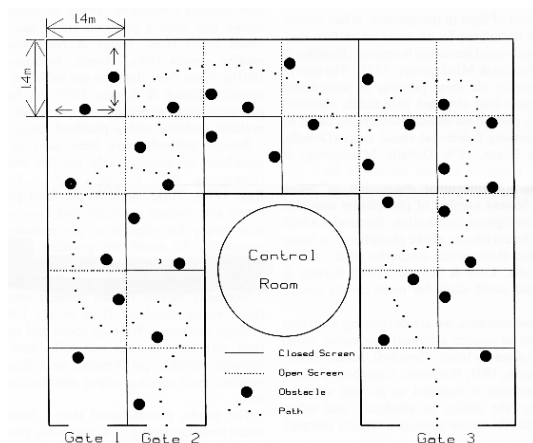


Figure 5.1 : le parcours utilisé par Cha et al. (1992). Des panneaux et des obstacles amovibles permettent de générer aléatoirement des parcours de difficultés et de longueur équivalentes.



Figure 5.2 : environnement virtuel utilisé par l'équipe de Dagnélie (2007)

l'apprentissage du dispositif. En effet, le temps nécessaire pour parvenir à un certain niveau de performance est également un critère à prendre en compte dans l'évaluation du dispositif. De plus, si l'on souhaite comparer les performances de deux dispositifs, il faut pouvoir s'assurer que les différents groupes ont bénéficié du même apprentissage.

Dans la plupart des études, les sujets suivent donc un entraînement systématique visant à leur permettre de réaliser le parcours de test. Un problème se pose alors : si l'on utilise le parcours de test pour entraîner le sujet, ce dernier est susceptible de le mémoriser, ce qui fausse la mesure de l'effet lié au dispositif. Néanmoins, les parcours d'entraînement doivent être plus ou moins similaires au parcours de test afin que le sujet puisse y acquérir des compétences utiles lors du test. L'obtention de compétences opérationnelles sur tout type d'environnement est en effet susceptible de nécessiter un entraînement très long, incompatible avec les contraintes expérimentales usuelles.

Dans ce but, Cha et ses collaborateurs (Cha *et al.*, 1992) utilisent un dispositif constitué de panneaux et d'obstacles amovibles permettant de réaliser des parcours aléatoires de taille et de difficultés équivalentes (fig. 5.1). Cette méthode nécessite toutefois une infrastructure lourde (pour maintenir les panneaux et les obstacles), et est vraisemblablement difficile à mettre en pratique en extérieur.

De son côté, Dagnélie (2007) propose de sélectionner, dans un environnement donné, plusieurs parcours structurellement équivalents : en l'occurrence des parcours de petite taille correspondant à une portion de couloir muni de portes. Dans ce cas, la difficulté consiste à trouver suffisamment de parcours équivalents dans un lieu donné. Reste enfin la possibilité de tester les sujets dans des environnements virtuels (fig. 5.2, tirée de Dagnélie, 2007). Dans ce cas toutefois, il est souvent difficile d'établir si les résultats seront généralisables en environnement réel. Par ailleurs, les déplacements du sujet dans l'environnement virtuel sont réalisés à partir d'un joystick, ce qui limite les interactions sensorimotrices possibles.

Pour notre part, nous faisons le constat suivant : la réalité quotidienne d'un sujet non-voyant consiste bel et bien à réaliser systématiquement le même parcours, par exemple pour se rendre de son domicile à son lieu de travail. Nous nous sommes donc penchés sur la possibilité d'utiliser un seul et même parcours pour l'apprentissage et pour l'évaluation du dispositif. La question qui se pose alors est : comment dissocier l'effet lié au dispositif de l'effet lié à la mémorisation du





Figure 5.3 : étape de familiarisation. L'opérateur guide le sujet le long du parcours en le tenant par le bras

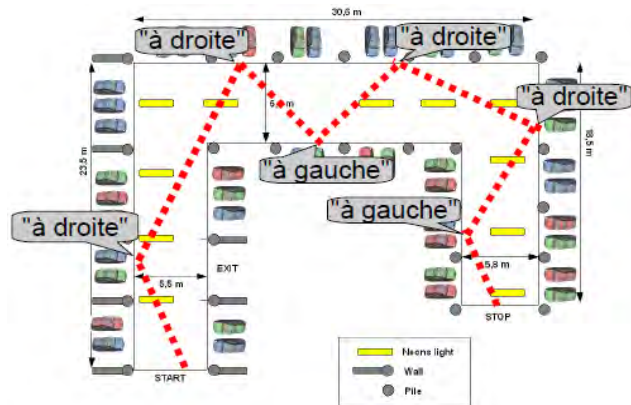


Figure 5.4 : étapes d'entraînement. Le sujet effectue le parcours seul. L'opérateur fournit une indication verbale lorsque le sujet sort du parcours.

parcours?

## 5.2.2 Méthode

La méthode d'évaluation que nous proposons utilise donc un seul et même parcours. Elle est séparée en deux phases : une phase d'apprentissage et une phase de test. Dans la phase d'apprentissage, le sujet, muni du dispositif de substitution, effectue plusieurs fois le parcours. La phase de test est, pour le sujet, totalement similaire aux phases d'apprentissages. A son insu néanmoins, nous perturbons le fonctionnement du dispositif, ce qui nous permet d'isoler très proprement son effet.

### Phase d'apprentissage :

La phase d'apprentissage est réalisée sur plusieurs sessions séparées d'au moins un jour.

Chaque session démarre par ce que nous appelons la familiarisation : le sujet, équipé du dispositif, réalise le parcours en étant guidé par l'opérateur qui lui tient le bras (fig. 5.3). Ce type d'apprentissage passif est uniquement destiné à familiariser le sujet à la tâche, ou à le replonger dans l'utilisation du dispositif.

Vient ensuite l'apprentissage à proprement parler. L'utilisateur, toujours muni du dispositif est invité à réaliser le parcours seul. L'opérateur, situé derrière lui, fournit des indications verbales lorsque le sujet atteint les bords du parcours. Les seules indications fournies sont "à gauche" ou "à droite" de manière à ramener le sujet vers le parcours (fig. 5.4). Le nombre d'indications verbales fournies est comptabilisé. Il constitue ce que nous appelons le "nombre de contacts" (NC). Le temps mis par le sujet pour réaliser le parcours est également mesuré (RT pour Run Time). Cette opération est répétée plusieurs fois, selon le protocole choisi.

### Phase de test :

La session de test est, du point de vue du sujet, totalement similaire à une session d'apprentissage. De manière à susciter une meilleure concentration, ce dernier est uniquement informé qu'il s'agit de la session de test. Tout comme dans les sessions d'apprentissage, après l'étape de familiarisation, le sujet est invité à réaliser plusieurs fois le parcours seul. Dans l'un des deux

derniers passages toutefois, l'image provenant de la caméra est inversée à l'insu du sujet (en opérant une symétrie par rapport à l'axe vertical). Comme dans les sessions d'apprentissage, on enregistre le nombre de contact et le temps de parcours du sujet dans les deux conditions. La position de l'essai où l'image est inversée est contrebalancée entre les sujets. Afin d'assurer un test en double aveugle, l'opérateur ne sait pas dans quel essai l'image est inversée.

### **Effet du dispositif :**

L'effet induit par le dispositif se mesure par la différence des performances du sujet entre la condition inversée et la condition normale. L'hypothèse est que les sujets sont meilleurs dans la condition "Normale", qui correspond au couplage sensori-moteur appris, que dans la condition "Inverse", où le couplage est différent du couplage appris.

Pour améliorer la sensibilité de la mesure, on définit la différence relative de temps de parcours (RRTD pour Relative Run Time Difference) et la différence relative de nombre de contacts (RNCD pour Relative Number of Contact Difference), qui correspondent aux mesures de performances ramenées à la performance moyenne du sujet :

$$RRTD = 2 \cdot \frac{RT(Inverse) - RT(Normal)}{RT(Inverse) + RT(Normal)}$$

$$RNCD = 2 \cdot \frac{NC(Inverse) - NC(Normal)}{NC(Inverse) + NC(Normal)}$$

A priori, aucun des indicateurs ne doit être significativement négatif : cela impliquerait que les utilisateurs sont meilleurs lorsqu'ils n'utilisent pas le couplage appris. Notre hypothèse se traduit de plus par le fait que l'un au moins des indicateurs doit être strictement positif. On doit donc avoir :

$$\{RRTD \geq 0 \text{ et } RNCD \geq 0\} \text{ et } \{RRTD > 0 \text{ ou } RNCD > 0\}$$

### **Remarque :**

Le lecteur s'interrogera peut être sur la pertinence d'utiliser, comme situation de comparaison par rapport à une utilisation du dispositif en fonctionnement normal, une condition où le sujet est muni du dispositif dans lequel l'image d'entrée est inversée, plutôt qu'une situation où le sujet n'est pas équipé du dispositif, par exemple.

L'origine de ce choix provient d'une interrogation sur l'effet de contexte lors du test : une situation où le sujet n'est pas équipé du dispositif est-elle équivalente, en tout autre point, à une situation où le sujet est équipé du dispositif ?

Nous pensons que non, et en particulier que le sujet, sans le dispositif, est susceptible d'utiliser des stratégies différentes que lorsqu'il en est équipé. Il peut en particulier se concentrer sur les autres informations en provenance du milieu extérieur : tactiles (courant d'airs ou bouches d'aérations), olfactives (par exemple les odeurs d'essences près d'un garage), et bien sûr sonores, puisqu'il n'est plus équipé du dispositif. Par ailleurs, s'il n'est pas équipé du dispositif, le sujet est dans une situation où seul l'expérimentateur est susceptible de lui fournir une information (autre que celles discutées précédemment). Il peut donc choisir de marcher plus vite, en attendant que l'opérateur lui fournisse des indications. Enfin, le fait de ne pas être munis du dispositif signale de manière évidente qu'il s'agit de la condition contrôle, ce qui peut altérer la motivation du sujet et le comportement de l'expérimentateur.

Pour cette raison, nous n'avons pas choisi de comparer une situation avec le dispositif avec une situation sans le dispositif. A ce titre, l'intérêt de notre méthode est qu'elle permet d'isoler un effet

induit par l'image, et seulement par l'image. Toutefois, pour s'assurer de la validité de ce propos, nous avons réalisé, dans l'expérimentation préliminaire, une mesure post-expérimentale des performances du sujet sans dispositif.

Une autre solution aurait été de fournir au sujet un "faux" signal, par exemple un signal enregistré lors de la passation d'un autre sujet. Dans ce cas néanmoins, le sujet s'aperçoit rapidement de l'absence de couplage sensori-moteur effectif (Auvray, 2005). Il est donc susceptible de se comporter comme s'il n'était pas muni du dispositif. Dans le cas où l'on inverse l'image, le couplage sensori-moteur reste effectif, même s'il ne correspond pas au couplage appris.

Enfin, cette méthode présente également un avantage majeur : elle permet de maximiser l'effet potentiel du dispositif. En effet, si l'on suppose que l'apprentissage du couplage sensori-moteur induit par le dispositif va améliorer les performances du sujet dans la tâche de mobilité, l'utilisation d'un couplage différent de celui appris est susceptible de perturber le sujet, de diminuer ses performances et donc d'augmenter d'autant l'effet mesuré. Cet élément est d'autant plus important que l'on s'attend, dans un premier temps et compte tenu du temps d'apprentissage réduit, à obtenir un effet petit.

Enfin, cette méthode permet de s'affranchir des différences inter-individuelles puisque l'on normalise les données par les performances globales du sujet. Les indicateurs ainsi utilisés sont donc *a priori* plus sensibles que les données brutes. Ils sont également comparables d'un individu à l'autre, et d'une expérience sur l'autre.

Cette méthode ne permet pas, en revanche, de tester l'utilité du dispositif dans la pratique. Ce n'est d'ailleurs pas son objet : cette méthode permet de tester l'établissement d'un couplage sensori-moteur favorable à la mobilité, ce qui est une condition préalable à l'utilité pratique du dispositif. Elle permet donc de déterminer si le dispositif est un candidat potentiel pour être une aide à la mobilité. Le cas échéant, les résultats obtenus seront un argument favorable au test du dispositif à plus longue échelle. L'effet mesuré reste par ailleurs un indicateur relatif de la fonctionnalité du dispositif : un effet fort signalera un couplage fort, un dispositif induisant un effet fort dans cette expérience sera donc *a priori* plus efficace qu'un dispositif présentant un effet faible.

## 5.3 Expérience 1 : Validation de la méthode et du dispositif TheVIBE

Le but de cette première expérimentation était de valider la méthode proposée précédemment en mesurant l'effet induit sur des sujets équipés du dispositif TheVIBE, présenté au chapitre 3.

Cette expérience préliminaire a été réalisée dans le parking du laboratoire, avenue Félix Viallet (fig. 5.5). Ce lieu présentait l'avantage, dans le cadre d'une expérience sur la mobilité, d'être couvert, de posséder un éclairage constant, une complexité raisonnable, d'être peu passant et de présenter des surfaces piétonnes larges et sans obstacles. Le parcours est constitué de trois segments formant un U d'une longueur d'environ 73 mètres (fig. 5.6). Les bords du parcours sont délimités par des pylônes en bétons et par les voitures garées dans les emplacements. L'éclairage provient de néons fixés au plafond.

### 5.3.1 Participants

Vingt quatre sujets voyants et sans problèmes auditifs avérés ont participé à l'expérience : douze femmes et douze hommes. Leur âge s'étend de 22 à 38 ans, avec une moyenne de 26 ans et un écart-



Figure 5.5 : le parking du laboratoire. Le parking du laboratoire offre un terrain large et sans obstacles, peu fréquenté, avec un éclairage constant.

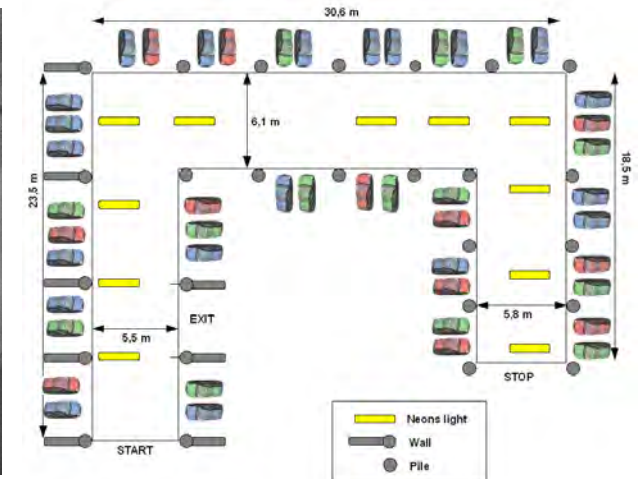


Figure 5.6 : schéma du parcours. Le parcours est constitué de trois segments formant un U d'une longueur de 73 mètres environ.

type de 5 ans. Seul trois d'entre eux ont déjà utilisé le dispositif TheVIBE, dans le cadre de l'expérience de localisation décrite au chapitre précédent.

### 5.3.2 Appareillage

Les sujets portent un masque semi-rigide doublé de disques en coton pour garantir une occultation complète et efficace du champ visuel. Ils sont équipés du dispositif TheVIBE comprenant :

- Une caméra frontale Blackeye2 présentant un champ de 92°
- Un casque stéréophonique Seinheiser HD 280 Pro dont la réponse en fréquence est particulièrement fidèle
- Un ordinateur portable Dell Latitude 620 équipé d'un processeur Intel Core 2 cadencé à 2 Ghz, et de 2 Go de mémoire vive

L'image en provenance de la caméra est transmise au PC via un convertisseur PCTV Pinnacle USB Pro. Cette image est traitée par le logiciel TheVIBE qui génère un son transmis au sujet par les écouteurs.

#### Configuration du logiciel TheVIBE :

Pour cette expérience, le PC est équipé de la version 2.1 de TheVIBE.

Cette version comprend en particulier :

- Le codage de l'activité des champs récepteurs par une modulation de l'intensité sonore associée selon une échelle linéaire en dB (voir Ch. 3 §3.3.1)
- Le codage du panoramique par différence interaurale d'intensité sonore selon une échelle linéaire en dB allant de -12dB à +12 dB (voir Ch. 3 §3.3.2)
- Le codage de la tonie selon une échelle linéaire en Barks, dans l'intervalle 300-3000 Hz, soit

approximativement 3-16 Barks (voir Ch. 3 §3.3.3).

Le fichier de configuration, unique pour chaque utilisateur, comprend 200 champs récepteurs dont les centres sont répartis grâce à un algorithme d'auto-organisation de type Kohonen (ch.3 §2.4). Chaque champ récepteur est composé de 10 échantillons distribués aléatoirement selon une distribution de probabilité gaussienne, centrée sur le centre du champ récepteur et d'écart-type 5 pixels.

Le "mapping" visuo-auditif est de type "Euclidien" (ch. 3 §3.5.1).

La distribution des échantillons dans l'image est corrigée pour tenir compte de l'optique de la caméra (ch. 3 §2.6). Pour palier un léger décentrage de l'optique de la caméra, qui induit une translation de l'image, la fenêtre de travail dans l'image d'origine a été réduite à 260 par 220 pixels.

### **Inversion de l'image dans la phase de test :**

Sur cette version de TheVIBE, la possibilité d'inverser l'image en provenance de la caméra ne peut se faire que manuellement, par l'opérateur. Pour que l'opérateur ne sache pas dans quel essai l'image est inversée, nous avons donc utilisé un autre moyen : nous avons utilisé un deuxième fichier de configuration (ou fichier "rétine") dans lequel la position des échantillons est "inversée", c'est à dire qu'ils ont subi une symétrie par rapport à l'axe des ordonnées, sans pour autant changer les propriétés sonores attribuées à chaque champ récepteur. Cette opération est totalement équivalente à une inversion gauche/droite de l'image.

On crée par la suite deux fichiers d'options avec des noms banalisés (option\_a.txt et option\_b.txt par exemple), l'un mentionnant la rétine utilisée pour l'apprentissage, et l'autre, la rétine inverse. L'opérateur manipule uniquement les fichiers d'option et ne sait donc pas quelle rétine est utilisée pour quel essai.

## **5.3.3 Protocole**

Le protocole expérimental comprend en tout 4 sessions : 3 sessions d'apprentissage et une session de test. Un jour d'écart au moins est laissé entre chaque sessions. La session d'apprentissage et la session de test sont séparées au maximum de 8 jours.

Aucune information concernant le dispositif n'est fournie au sujet, à l'exception du fait que le son entendu dans les écouteurs est le fruit d'un codage de l'image en provenance de la caméra. En particulier, aucune indication sur le codage visuo-sonore n'est fournie au sujet.

Chaque session d'apprentissage est composée de 3 passages dans le parcours : une première fois durant laquelle le sujet est tenu par l'opérateur au niveau du bras (essai de familiarisation), puis deux fois durant lesquelles le sujet est libre de ses mouvements et est guidé verbalement par l'opérateur (essais d'apprentissage). L'information fournie par l'opérateur se limite à deux possibilités : "à gauche", ou "à droite". Elle est délivrée uniquement lorsque le sujet atteint les bords du parcours, et de manière à le ramener vers le parcours. Le nombre d'indications verbales fournies ainsi que le temps mis pour réaliser le parcours sont enregistrés pour chaque essai.

La session de test est, du point de vue du sujet, entièrement similaire aux sessions d'apprentissage. Elle est composée de 3 passages, le premier conduit par l'opérateur et les deux suivantes en guidage verbal. A l'insu du sujet toutefois, l'image est inversée dans l'un de ces deux derniers essais. Le nombre d'indications verbales fournies ainsi que le temps mis pour réaliser le parcours sont enregistrées pour chaque essai.

A l'issu de ce protocole, à titre de contrôle et hors du cadre strict de l'expérimentation, nous avons effectué deux essais supplémentaires où le sujet n'est pas équipé du dispositif (mais garde les yeux bandés). Le nombre d'indications verbales fournies ainsi que le temps mis pour réaliser le parcours ont également été enregistrés.

### 5.3.4 Résultats

Le tableau 1 présente les mesures effectuées lors de cette expérience au cours des trois sessions d'apprentissage (moyennées sur les deux essais d'apprentissage) et au cours de la session de test. Quatre sujets n'ont pas été inclus dans l'analyse des données car il ne respectaient pas les conditions du protocole expérimental : le temps entre leur dernière session d'apprentissage et leur session de test était supérieur à 8 jours.

Pour les sessions d'apprentissage, les données considérées sont le temps mis pour réaliser le parcours (respectivement RT1, RT2 et RT3), et le nombre de fois où le sujet est sorti du parcours (respectivement NC1, NC2 et NC3).

Pour la session de test, les données consignées sont les temps mis pour réaliser le parcours dans l'essai "normal" et dans l'essai "inversé" (respectivement RT\_Norm et RT\_Inv), et le nombre de contacts dans l'essai "normal" et dans l'essai "inversé" (respectivement NC\_Norm et NC\_Inv).

Tableau 1 : Données issues de l'expérience de validation en situation de mobilité

Sujet	RT1	NC1	RT2	NC2	RT3	NC3	RT_Inv	NC_Inv	RT_Norm	NC_Norm
2	173	8.5	173	5	162	6.5	176	6	210	8
3	168	6.5	131	5.5	139	6	161	5	110	4
5	160	7.5	182	8	254	9.5	230	6	207	4
6	206	6	158	5	102	4	151	4	156	7
7	237	7	181	7	129	6	113	5	115	4
8	592	9	502	8	298	6.5	311	8	347	6
9	339	6.5	150	7	178	5.5	221	6	144	5
11	248	7	187	5	170	3.5	181	5	124	2
12	334	9	247	5	256	7	285	6	258	9
13	211	6.5	193	6	147	5	132	8	129	3
15	264	9.5	168	6.5	154	4	114	5	130	4
16	145	6	107	4	104	4.5	103	7	117	7
17	307	11	373	14	331	7.5	399	11	340	8
18	284	7	213	3.5	220	6.5	214	7	197	7
19	168	6	121	5.5	115	2.5	104	4	118	6
20	137	7	132	6.5	111	6.5	161	8	88	5
21	152	8	109	5	124	10	102	8	97	3
22	112	2.5	97	4.5	118	8.5	86	5	85	3
23	322	7.5	342	10	198	7.5	296	12	425	18
24	139	8	164	6	132	2	202	7	150	4
Moyenne	234,63	7,3	196,25	6,35	171,88	5,95	187,1	6,65	177,35	5,85

#### 5.3.4.1 Analyse des sessions d'apprentissage

La figure 5.7 présente les mesures de performance des sujets (RT et NC) en fonction de la session d'apprentissage

Concernant l'apprentissage, nous souhaitons bien sûr vérifier qu'il y a bien eu un effet de la

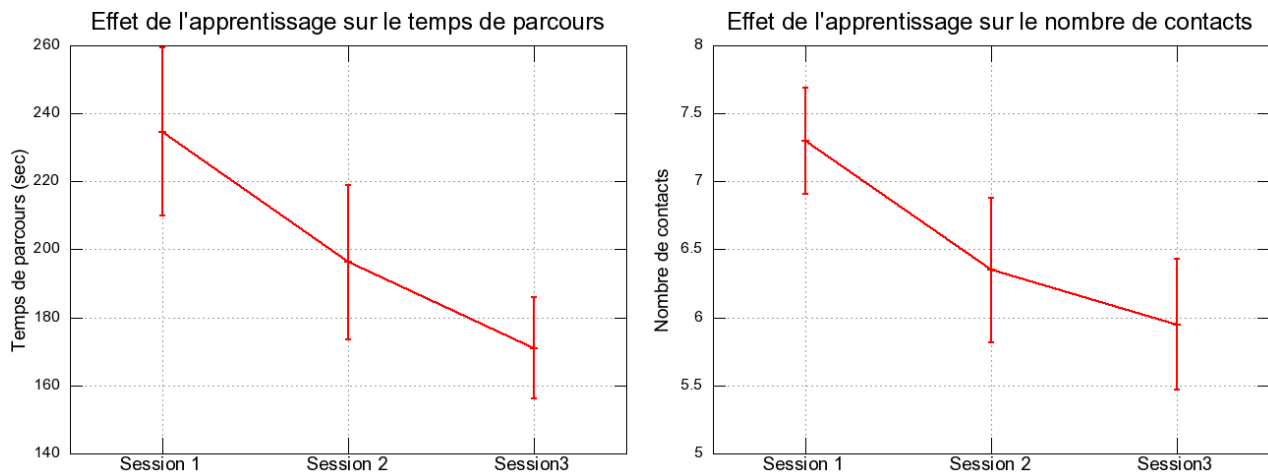


Figure 5.7 : effet de l'apprentissage sur les performances du sujet.

session d'apprentissage mais nous voudrions aussi savoir si cet apprentissage était quantitativement suffisant ou si les performances des sujets auraient pu encore s'améliorer de beaucoup.

Pour cela, nous souhaitons effectuer un test statistique sur deux contrastes, autrement dit sur deux comparaisons. La première hypothèse est qu'il y a une différence de performance entre la première session et les deux sessions suivantes. La deuxième hypothèse est qu'il n'y a pas de différence significative entre la deuxième et la troisième session. Pour chaque indicateur, nous avons donc deux contrastes orthogonaux:

Contraste n° 1 :  $2 * \text{Session 1} - \text{Session 2} - \text{Session 3}$

Contraste n°2 :  $\text{Session 2} - \text{Session 3}$

Pour tester ces contrastes à l'aide d'un test de Student, nous devons vérifier que leur distributions ne diffèrent pas significativement d'une distribution Gaussienne. Nous utilisons pour cela un test de Shapiro et Wilk (1965). Pour les données concernant le nombre de contacts, la distribution des données issues des deux contrastes ne diffère pas significativement d'une distribution gaussienne ( $p = 0.29$  pour le contraste 1, et  $p = 0.96$  pour le contraste 2)

Pour les données sur le temps de parcours en revanche, elles ne le sont pas ( $p < .01$  pour le deuxième contraste). Pour pallier à cela, nous avons appliqué une transformation logarithmique sur les données de temps de parcours. La distribution des contrastes sur le logarithme des temps de parcours n'est alors plus significativement différent d'un modèle gaussien ( $p = 0.89$  pour le contraste 1 et  $p = 0.83$  pour le contraste 2). Nous avons ensuite effectué un test de Student pour chaque contraste et chaque indicateur. Les résultats sont présentés dans le tableau 2.

Tableau 2 : tests statistiques des contrastes sur les sessions d'apprentissages

	Temps de parcours : log(RT)	Nombre de Contacts : NC
<b>Contraste 1</b>	Significatif ( $p < .001$ )	Significatif ( $p < .05$ )
<b>Contraste 2</b>	Non significatif ( $p > .05$ )	Non significatif ( $p > .1$ )

Nos hypothèses sont donc bien vérifiées : les sujets s'améliorent significativement entre la première session et les deux suivantes. En revanche, les performances ne s'améliorent pas

significativement entre la deuxième session et la troisième session.

#### Remarque :

Nous montrons ici que les performances du sujet semblent stagner dès la deuxième session, ceci implique-t-il que l'apprentissage du dispositif est terminé?

La réponse est non. En effet, les mesures de temps de parcours et de nombre de contacts n'isolent pas l'effet lié au dispositif, elles comprennent d'autres effet, en particulier l'effet lié à l'apprentissage par cœur du parcours. Ce test n'est donc pas suffisant pour affirmer que l'apprentissage est terminé. Il donne en revanche une indication à ce sujet, car on peut dire qu'à partir de la deuxième session, les mesures directes de temps de parcours et de nombre de contacts ne sont plus assez sensibles pour déceler un effet du dispositif.

#### 5.3.4.2 Analyse de la session de test

A partir des mesures de temps de parcours et de nombre de contacts dans chacune des conditions (Image inversée, image normale), on calcule les différences relatives de performances associées au temps de parcours (RRTD) et au nombre de contacts (RNCD).

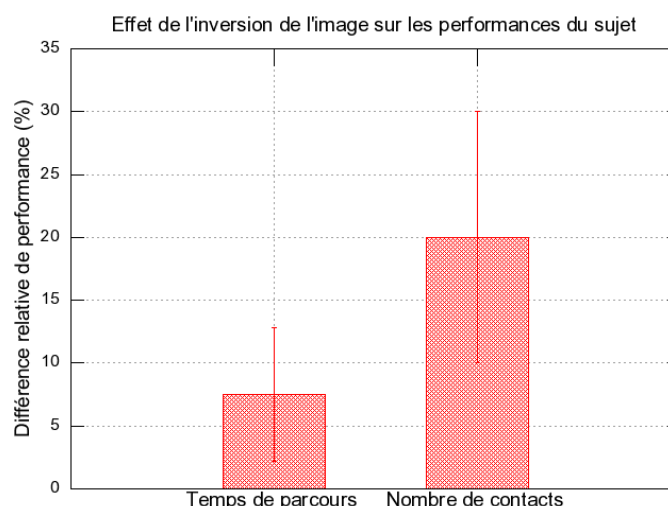


Figure 5.8 : effet du dispositif sur les temps de parcours et sur le nombre de contacts

La figure 5.8 présente ces différences relatives de performances pour les temps de parcours et pour le nombre de contacts.

Si les sujets ont bien appris le couplage sensori-moteur induit par le dispositif, et si ce couplage a un effet sur leur comportement, leurs performances dans la condition "Image normale" (couplage appris) devraient être meilleures que dans la condition "Image inversée" (couplage non-appris). Nous devons donc tester si l'un au moins des deux indicateurs (RRTD ou RNCD) est significativement supérieur à 0. Le test d'hypothèse est donc unilatéral.

Les distributions des valeurs de RRTD et de RNCD ne sont pas significativement différentes d'une distribution gaussienne ( Test de Shapiro et Wilk,  $p = 0.54$  pour le RRTD ,  $p = 0.27$  pour le RNCD), on peut donc utiliser un test de Student. Les résultats de ce test sont présentés dans le tableau 3. La différence relative du nombre de contact est donc significativement supérieure à 0. L'inversion de l'image a donc bien un effet sur les performances du sujet.



Tableau 3: tests statistiques des différences relatives de performances dans la session de test

	<b>RRTD</b>	<b>RNCD</b>
<b>Test de Student unilatéral (Moyennes &gt; 0 )</b>	Non significatif (p = 0.08)	Significatif (p < .05)

### 5.3.5 Discussion

Nous avons donc pu mettre en évidence un effet du dispositif de substitution sur la mobilité des sujets. Cet effet concerne le nombre d'erreurs de contacts, c'est à dire le nombre de fois où le sujet est sorti du parcours. En ce sens, le dispositif a donc un effet sur la trajectoire du sujet.

#### **Analyse statistique des données :**

La différence observée sur le temps de parcours n'est, en revanche, pas significative. Il est toutefois vraisemblable qu'elle le devienne si l'on augmente le nombre de sujets. Une question se pose par ailleurs quand aux mesures de temps de parcours. Nous avons vu dans le cas des sessions d'apprentissages que la distribution des temps de parcours ne pouvait pas être considérée comme gaussienne, mais que la distribution du logarithme des temps de parcours l'était. On pourrait donc envisager d'utiliser le logarithme des temps de parcours pour calculer les différences relatives de performance. Nous avons donc réalisé ce test, uniquement à titre indicatif. La distribution des différences relative des logarithmes des temps de parcours n'est pas significativement différente au sens strict d'une distribution gaussienne mais le p du test de Shapiro-Wilk est très faible (p = 0.055). En revanche, l'indicateur est cette fois, significatif (p < .05). Cette question du traitement des données temporelles est donc pertinente.

Nous avons, par ailleurs, regardé à titre indicatif les résultats obtenus avec les données brutes, c'est à dire sans normaliser les performances par les performances moyennes du sujet. Sur les temps de parcours, cette normalisation permet de rendre les distributions "gaussiennes". Sur les nombres de contacts, cette normalisation améliore effectivement la sensibilité de la mesure.

#### **Mesures post expérimentales sans le dispositif :**

Comme nous l'avons mentionné dans le protocole, nous avons effectué, une fois l'expérience terminée, une mesure des performances des sujets sans le dispositif. Nous avons ensuite comparé les performances des sujets avec et sans le dispositif en calculant, de la même manière que précédemment, la différence relative entre les performances avec le dispositif et les performances sans le dispositif. La figure 5.9 illustre les résultats obtenus.

Les distributions des différences relatives de temps de parcours ne diffèrent pas significativement d'une distribution gaussienne, nous avons donc utilisé, pour tester la significativité de nos mesures, un test de Student. Cette fois, cependant, le test est bilatéral (tableau 4).

Tableau 4 : test statistique des différences relatives de performances avec et sans le dispositif.

	<b>Temps de parcours</b>	<b>Nombre de Contacts</b>
<b>Test de Student bilatéral</b>	Non significatif (p = 0.05016)	Significatif (p < .001)

Les sujets semblent donc avoir une meilleure trajectoire avec le dispositif que sans. Les résultats concernant le temps de parcours sont non significatifs mais avec un p très proche de 5%. Sous réserve de confirmation, il semble donc que les sujets soient plus rapide sans le dispositif qu'avec.

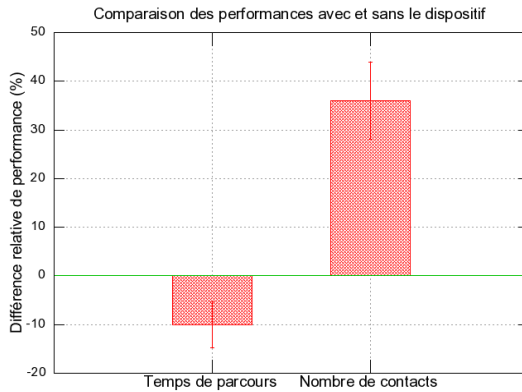


Figure 5.9 : Comparaison des performances des sujets avec et sans le dispositif

De tels résultats ne permettraient pas de conclure quant à l'effet du dispositif sur les performances des sujets : notre hypothèse concernant l'existence de différentes stratégies semble donc fondée et le choix de comparer deux situations où le sujet est équipé du dispositif justifié.

### 5.3.6 Conclusion

Cette première expérience a bien permis de mettre en évidence l'apprentissage d'un couplage sensorimoteur par le sujet, et le fait que cet apprentissage a un effet sur la mobilité du sujet. Le dispositif est donc un candidat potentiel pour une aide à la mobilité.

L'effet du dispositif reste toutefois relativement faible (de l'ordre de 20% de différence pour le nombre de contacts). En pratique les différences de performances étaient peu visibles. Il n'était pas possible pour l'opérateur de deviner quel était l'essai "inversé".

On peut proposer plusieurs explications à cela. Tout d'abord, le temps d'apprentissage était quantitativement faible (de l'ordre de 20 minutes par session soit 1h en tout). Cela dit, nous avons vu que les performances du sujet stagnaient entre la deuxième et la troisième session, ce qui semble suggérer que l'apprentissage était suffisant. Reste néanmoins à vérifier qu'il ne s'agissait pas d'un phénomène temporaire, c'est à dire d'un simple palier d'apprentissage.

Une autre possibilité est que l'environnement choisi soit trop simple. Une remarque fréquente est qu'en dehors des deux virages, la structure de l'environnement reste relativement invariante par une symétrie selon l'axe vertical (figure 5.10), ce qui peut réduire l'effet lié à une inversion de l'image de la caméra. Cette remarque est toutefois à prendre avec précaution. Elle est une belle illustration des différences de raisonnement que l'on peut tenir suivant que l'on considère une approche passive ou une approche sensori-motrice de la perception. En effet, si du point de vue passif l'image inversée ressemble beaucoup à l'image normale, il n'en est pas moins vrai que, du point de vue actif, les contingences sensori-motrices du dispositif sont complètement inversées. Considérons par exemple la trace lumineuse au sol, très similaire dans les deux images et qui peut vraisemblablement servir de point de repère pour les sujets. Dans la condition "normale", si le sujet tourne sa tête vers la droite, cette ligne se décalera vers la gauche dans l'image. Dans la condition "inverse", elle se décalera vers la droite. Du point de vue des contingences sensori-motrices apprises, cette ligne directrice devient donc un objet en mouvement!

Néanmoins, la question de la simplicité de l'environnement est intéressante. D'un côté on peut supposer qu'il est plus facile d'apprendre à se servir d'un dispositif dans un environnement simple, car on va être plus à même d'interpréter la stimulation (approche passive), ou d'apprendre les contingences sensori-motrices (approche active). D'un autre côté, un environnement simple offre peu d'information, donc peu d'indices d'orientation (approche classique), ou un couplage sensori-



Figure 5.10 : Effet de l'inversion de l'image. (a gauche image normale, à droite, image inversée) En dehors des virages, l'inversion de l'image ne perturbe que peu la structure du parcours

moteur relativement pauvre (approche active) et favorisera donc un effet petit. Si l'on admet que dans notre cas l'apprentissage était suffisant, et qu'on se place dans l'approche active, on en déduit que la simplicité de l'environnement est dans ce sens susceptible de réduire l'effet du dispositif.

La dernière explication possible est enfin que le dispositif n'est pas efficace dans cette situation. C'est sur cette dernière hypothèse que nous allons nous pencher. Pour cela, nous allons maintenant voir comment nous avons utilisé ce même protocole pour tester l'intérêt d'un prétraitement biomimétique de l'image : la simulation du traitement du signal dans la voie Magno.

## 5.4 Expérience 2 : effet d'un traitement spatio-temporel de l'image sur l'efficacité du dispositif de substitution

Comme nous l'avons vu au chapitre 2, le traitement du signal dès la rétine est un élément clé du processus visuel entier. Le comportement spatio-temporel de la rétine est notamment à l'origine de la génération des voies Parvo et Magno, dont les propriétés fonctionnelles sont bien distinctes. C'est ici à la voie Magno que nous nous sommes intéressés puisqu'elle semble être plus particulièrement dédiée à l'information de contexte et de mouvement, et donc particulièrement utile pour la mobilité (voir chap. 2).

L'idée est donc ici de tester l'effet de ce "filtrage" spatio-temporel biomimétique lorsque ce dernier est appliqué à l'image de la caméra avant son entrée dans le dispositif de substitution.

Nous avons par ailleurs choisi un nouvel environnement de test : nous nous sommes placé en extérieur, dans les allées de la cour intérieure des bâtiments universitaires de l'IMAG (Institut des Mathématiques Appliquées de Grenoble, fig. 5.11 et 5.12). Le parcours choisi est long d'environ 65 m. Il comporte trois segments avec un virage à droite à  $45^\circ$  et un virage à gauche à  $45^\circ$ .



Figure 5.11 : vue de la cour de l'IMAG. L'environnement est plus complexe que dans le garage de l'INPG.



Figure 5.12 : le parcours extérieur dans la cour de l'IMAG

Le choix d'un environnement extérieur présente à notre sens plusieurs intérêts :

1 – Les systèmes naturels dont nous nous inspirons sont vraisemblablement adaptés à ce type d'environnements extérieurs. On peut donc légitimement supposer qu'il sera plus facile de mettre en évidence l'intérêt de traitements biomimétiques sur ce type de terrain.

2 – Un environnement extérieur présente une échelle de taille plus étendue qu'un environnement intérieur. En effet, dans un environnement intérieur, l'échelle de taille maximale correspond aux dimensions de l'espace de travail (de l'ordre de la dizaine de mètres dans le garage). En extérieur, cette dimension n'est pas limitée : au delà du parcours, il y a les arbres et les bancs avoisinants, puis au delà les bâtiments, puis plus loin les montagnes, et enfin le ciel et le sol. Cet aspect est d'autant plus crucial qu'il est très difficile de déterminer quelles échelles de taille sont susceptibles d'être utiles pour la tâche, et quelles échelles de tailles sont accessibles via le dispositif.

3 – Un environnement intérieur est souvent plus simple qu'un environnement extérieur, de ce fait la stimulation issue du dispositif peut vraisemblablement être analysée sémantiquement par le sujet. Dans le garage en particulier, les néons au plafond produisaient un son aisément identifiable. Plusieurs des sujets ont interprété ce son comme un dispositif de guidage : ils tâchaient de maintenir la direction apparente du son droit devant eux. Dans un environnement extérieur, ce type de raisonnement n'est plus possible car il n'y a plus d'indices caractéristiques aussi évidents. Les sujets sont donc plus naturellement amenés à ne pas chercher à "comprendre" la stimulation. L'environnement est par ailleurs en ce sens plus écologique : il n'y a pas d'indices directeurs artificiels susceptibles de guider le sujet.

## 5.4.1 Participants

Seize sujets voyants et sans problèmes auditifs avérés ont participé à l'expérience. Leur âge s'étend de 21 à 52 ans, avec une moyenne de 29 ans et un écart-type de 9 ans. Les sujets sont répartis en deux groupes appariés selon leur sexe et selon leur âge (test de Student,  $p = 0.54$ ). Par ailleurs, les groupes ont aussi été appariés de manière à ce que l'ordre de passation Image Normale/

Image Inversée dans la phase de test soit contrebalancée de la même manière dans tous les groupes, c'est à dire que tous les groupes présentent le même nombre de sujets ayant passé la condition "inverse" puis la condition "normale", et donc par extension le même nombre de sujets ayant passé la condition "normale" puis "inverse". Enfin, aucun des sujets n'avait déjà utilisé le dispositif TheVIBE.

## 5.4.2 Appareillage

L'appareillage des sujets est identique à l'expérimentation précédente, à l'exception du logiciel TheVIBE. Nous avons en effet développé une nouvelle version du logiciel (3.1) dans laquelle l'image d'entrée peut être traitée avant d'être convertie en son.

### Traitement de l'image

Afin de tester l'effet d'un filtrage spatio-temporel de l'image, nous avons séparé les sujets en deux groupes, l'un pour lequel l'image d'entrée n'est pas traitée (groupe "Contrôle"), l'autre où l'image d'entrée est traitée au préalable (groupe "Magno").

Le traitement appliqué à l'image dans le groupe Magno correspond à la simulation du comportement spatio-temporel des ganglionnaires Parasol de type ON. Le processus de traitement est schématisé figure 5.13. Pour des raisons de rapidité de calcul, le modèle ne considère pas le filtrage initial opéré par les photorécepteurs. L'image d'entrée subit un filtrage spatio-temporel passe-bas correspondant à la réponse des cellules horizontales. Les caractéristiques spatiales et temporelles du filtre, respectivement " $k$ " et " $\tau$ ", ainsi que le gain du filtre, " $\beta$ " sont paramétrables via le fichier d'option (voir chapitre 2).

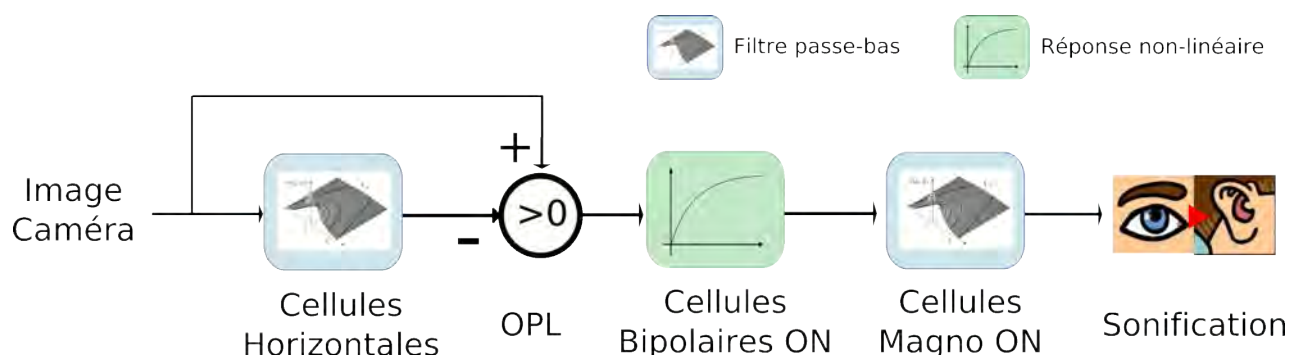


Figure 5.13 : schéma du filtrage implémenté dans le dispositif TheVIBE

La sortie des cellules bipolaires ON est calculée en prenant la partie positive de la différence entre l'image d'entrée et l'image filtrée. La réponse non linéaire des cellules bipolaires, de type Naka et Rushton, est ensuite simulée pour rehausser le signal. Cette non-linéarité n'est toutefois pas locale, elle est pilotée par un seul paramètre global " $X0$ ", déterminé par l'utilisateur dans le fichier d'options.

Pour être complet, il convient maintenant de considérer le filtrage passe-bas correspondant à l'échantillonnage de l'image dans le dispositif (voir chap. 3, partie 3). Ce filtrage peut être assimilé au filtrage passe-bas opéré par les cellules ganglionnaires parasol, dont nous avons vu qu'il conduisait naturellement aux propriétés spatio-temporelles des cellules de la voie Magno, en particulier à leur réponse transitoire. La voie simulée en prenant en compte l'échantillonnage de l'image est donc la voie Magnocellulaire ON.



Figure 5.14 : illustration du filtrage appliqué à l'entrée du dispositif TheVIBE (à droite) sur une image du parcours (à gauche).

Pour cette expérimentation, après quelques essais préliminaires, nous avons choisi les paramètres du traitement suivant :

- $k = 6$  pixels
- $\tau = 2$  images
- $\beta = 0.2$
- $X0 = 32$

Par ailleurs, l'échantillonnage de l'image dans cette expérimentation a été réalisé en utilisant la méthode d'auto-organisation de Kohonen. L'image traitée subit de ce fait en dernière étape un filtrage passe-bas gaussien. La taille de l'image est de 320 par 240 pixels. Le nombre de champs récepteurs est fixé à 200. La taille des champs récepteurs est fixée à 5 pixels. La fréquence de coupure du filtrage passe-bas est donc d'environ  $0,04 \text{ pixels}^{-1}$  (voir chapitre 3).

Par ailleurs, le taux de rafraichissement de la caméra est de 25 images par secondes. Le paramètre temporel  $\tau$  correspond donc à un temps caractéristique de l'ordre du dixième de seconde.

La figure 5.14 illustre ce filtrage sur une image.

**Remarque:**

Afin de garantir un test en double aveugle de l'effet du filtrage, l'activation du filtre se fait directement dans le fichier d'option. L'image de contrôle est par ailleurs banalisée de sorte que l'opérateur ne sait pas à quel groupe appartient le sujet.

### 5.4.3 Protocole

Le protocole est également similaire au protocole proposé pour l'expérience préliminaire. Chaque session comprend une étape de familiarisation. Compte tenu de la plus grande complexité de l'environnement, le nombre de passages d'apprentissage a toutefois été porté à 3 par session, au lieu de 2 précédemment. Dans la session de test, l'essai dans lequel l'image est inversée a lieu dans l'un des deux derniers passages.

## 5.4.4 Résultats

Le tableau 5 présente les mesures effectuées lors de cette expérience au cours des trois sessions d'apprentissages (moyenne sur les trois passages) et au cours de la session de test.

Tableau 5 : données issues de l'expérience sur l'effet du filtrage biomimétique en situation de mobilité

Sujet	Groupe	RT1	NC1	RT2	NC2	RT3	NC3	RT_Inv	NC_Inv	RT_Norm	NC_Norm
3	contrôle	278	9	256	4,67	212	5,67	233	8	227	10
5	contrôle	168	9,33	136	7,33	132	9,33	96	3	92	4
15	contrôle	154	9,67	242	12	266	10,67	275	10	289	13
16	contrôle	182	9,67	159	8	128	9	117	8	112	7
2	contrôle	170	11,67	142	8,67	108	5,67	149	13	142	8
8	contrôle	129	12	137	7	140	7	135	9	131	9
10	contrôle	182	7,67	183	3,33	178	4,67	153	5	144	4
12	contrôle	111	10	110	13,67	93	9	82	9	102	9
1	magno	99	5,33	101	7	105	6,33	102	5	111	6
6	magno	336	13,33	224	11	183	8,33	220	12	231	5
11	magno	124	7,67	116	6,33	97	3,33	86	7	81	3
14	magno	190	13,67	112	6	109	7	101	14	82	5
4	magno	164	7,67	154	11	138	9,33	107	8	96	10
7	magno	145	9	109	7	106	5,67	99	12	97	7
9	magno	147	10,67	113	11,67	98	10,67	75	5	74	7
13	magno	273	10	304	7,67	336	8,67	345	9	304	6
moyenne		178,21	9,77	162,38	8,27	151,9	7,52	148,44	8,56	144,69	7,06

### 5.4.4.1 Analyse des sessions d'apprentissage

La figure 5.15 présente les mesures de performance des sujets (RT et NC) en fonction de la session d'apprentissage, pour le groupe "Magno" et pour le groupe contrôle. Effet principal de la session

Pour effectuer une analyse statistique de l'effet de la session sur les performances du sujet, nous avons utilisé les contrastes définis dans l'expérience préliminaire : C1 (2,-1,-1), et C2 (0,1,-1). Tout comme précédemment, nous avons appliqué un logarithme aux données de temps de parcours avant de calculer le contraste. Nous avons ensuite effectué un test de Shapiro-Wilk pour tester la normalité des distributions des contrastes. Aucune des distributions des contrastes n'est significativement différente d'une distribution gaussienne (test de Shapiro-Wilk,  $p > 0.1$ ). Nous utilisons enfin un test de Student sur chaque contraste. Les résultats sont reportés dans le tableau 6.

Tableau 6 : test des contrastes sur les séances d'apprentissage

Mesure	Temps de parcours		Nombre de contacts	
Contraste	C1	C2	C1	C2
Test de Student	Significatif ( $p < 0.05$ )	Significatif ( $p < 0.05$ )	Significatif ( $p < 0.05$ )	Non Significatif ( $p > 0.1$ )



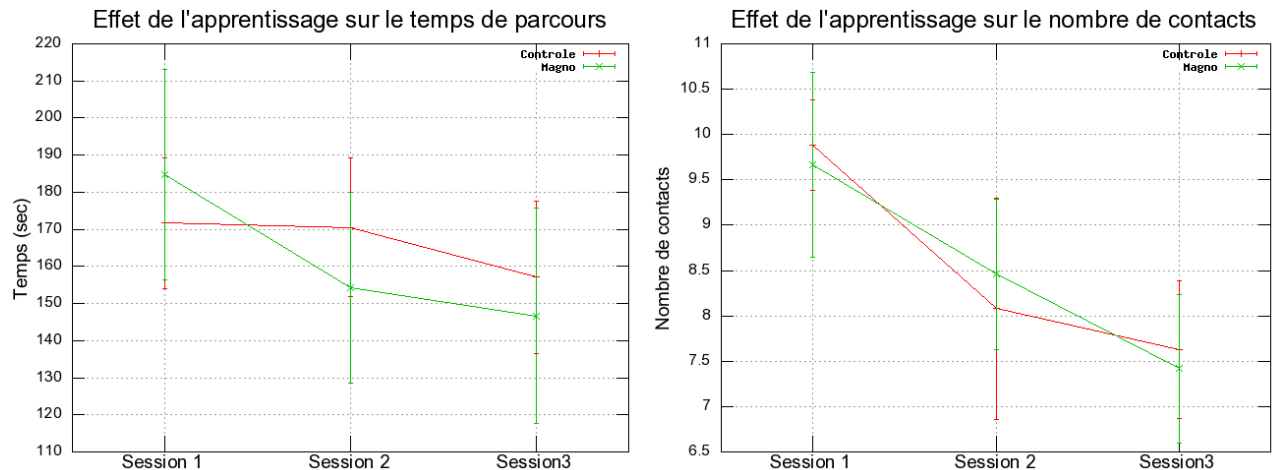


Figure 5.15 : effet de l'apprentissage sur les performances du sujet, pour le groupe filtré (tirets) et pour le groupe contrôle (croix)

### Effet principal du filtrage

Pour tester l'effet principal du filtrage, nous avons effectué la moyenne des performances des sujets sur les trois sessions d'apprentissage.

Nous avons ensuite vérifié la normalité des distributions des moyennes de temps de parcours et de nombre de contacts (test de Shapiro-Wilk,  $p > 0.05$ ). Nous avons vérifié que la variance de ces distributions n'étaient pas significativement différentes entre le groupe Magno et le groupe contrôle (test de Bartlett,  $p > 0.1$ ). Nous avons enfin effectué un test de Student pour tester l'effet du filtrage. Aucun effet du filtrage n'a pu être mis en évidence (test de Student,  $p > 0.1$  pour le temps de parcours et pour le nombre de contacts)

### Interaction entre l'effet de l'apprentissage et l'effet du filtrage

Pour tester l'interaction entre le filtrage et les contrastes d'apprentissage, nous vérifions que les distributions pour chaque couple (contraste, groupe) ne sont pas significativement différentes de distributions normales (Test de Shapiro-Wilk,  $p > 0.1$ ).

Nous vérifions ensuite que la variance des distributions des contrastes ne sont pas significativement différentes dans le groupe Magno et dans le groupe Contrôle (test de Bartlett,  $p > 0.1$  pour tous les contrastes).

Nous utilisons donc un test de Student. Il s'avère qu'aucune interaction entre le filtrage et l'apprentissage ne peut être mise en évidence (test de Student,  $p > 0.1$  pour tous les contrastes)

#### 5.4.4.2 Analyse de la session de test

La figure 5.16 présente les mesures de performances relatives sur le temps de parcours et sur le nombre de contacts pour le groupe Magno et pour le groupe contrôle.

Pour l'analyse statistique, nous avons choisi de tester l'effet du dispositif dans chacun des groupes Magno et contrôle, puis l'interaction éventuelle entre le filtrage et l'effet du dispositif.



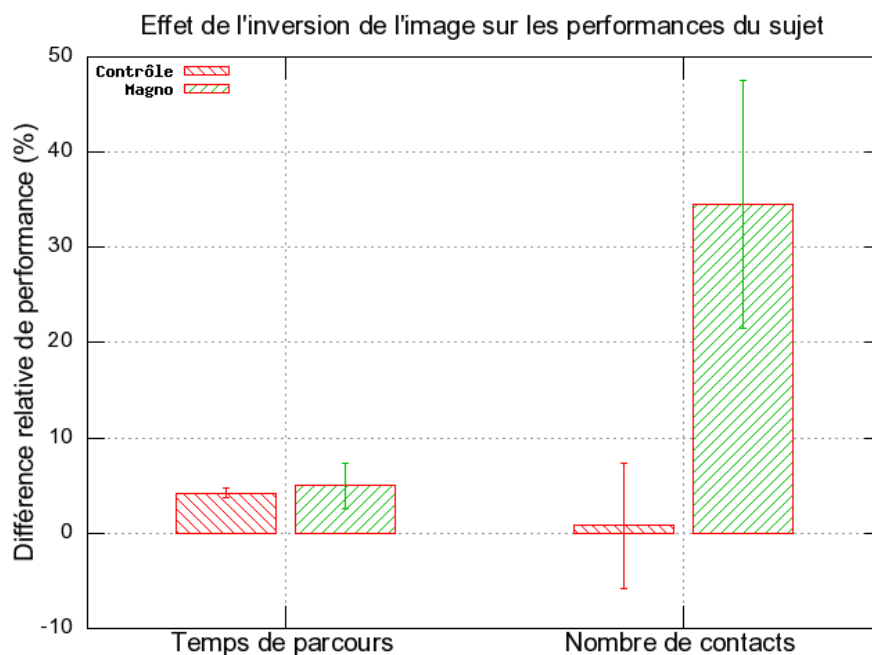


Figure 5.16 : Effet du dispositif sur les temps de parcours et sur le nombre de contacts en fonction du groupe.

### Effets simples du dispositif

Pour analyser statistiquement l'effet du dispositif, nous avons préalablement testé la normalité des indicateurs de performance (RRTD et RNCD) dans chacun des groupes Magno et contrôle. Toutes les distributions sont compatibles avec une loi normale (test de Shapiro-Wilk,  $p > 0.1$ ), à l'exception de la distribution des différences relatives de temps de parcours dans le groupe contrôle. Dans ce groupe nous avons identifié deux outliers (test de Grubbs,  $p < 0.01$ ) : il s'agit des sujets 12 et 15. Sans ces outliers, la distribution n'est plus significativement différente d'une distribution gaussienne (test de Shapiro-Wilk,  $p > 0.1$ ). Nous utilisons ensuite un test de Student pour déterminer si les distributions sont significativement positives (au sens strict). Tout comme précédemment, on utilise un test unilatéral. Les résultats sont présentés dans le tableau 7.

Tableau 7 : Test de de la positivité stricte des différences relatives de temps de parcours et de nombre de contacts dans chaque groupe

Groupe contrôle	RRTD	Significatif ( $p < 0.001$ )
	RNCD	Non significatif ( $p > 0.1$ )
Groupe Magno	RRTD	Non significatif ( $p = 0.09$ )
	RNCD	Significatif ( $p < 0.05$ )

On observe un effet significatif du dispositif dans le groupe contrôle ainsi que dans le groupe Magno.

### Interaction du filtrage avec les différences relatives de performances

Pour tester l'interaction entre le filtrage et les différences relatives de performances, on compare maintenant les distributions de différences relatives de temps de parcours et de nombre de contact entre le groupe "magno" et le groupe contrôle.

La normalité des distributions concernées a été testée dans le paragraphe précédent. L'homogénéité des variances est testée à l'aide d'un test de Bartlett ( $p > 0.05$  pour les deux indicateurs). On utilise donc un test de Student pour comparer les distributions. Ces tests concluent que les différences de temps de parcours relatif et de nombre relatif de contacts entre le groupe contrôle et le groupe Magno ne sont pas significatives (test de Student,  $p > 0.1$ )

## 5.4.5 Discussion

Dans cette expérimentation, nous avons testé la capacité de sujets à réaliser un parcours en extérieur à l'aide du dispositif TheVIBE. Les sujets étaient des sujets voyants avec les yeux bandés. Aucun d'entre eux n'avait utilisé le dispositif préalablement. Aucune indication concernant le fonctionnement du dispositif ne leur avait été fournie.

Les sujets étaient répartis en deux groupes. Un groupe où le comportement spatio-temporel des cellules Magno ON était simulé sur l'image de la caméra (groupe Magno) et un groupe contrôle.

Le test se déroulait sur quatre sessions. Les trois premières étaient dédiées à l'apprentissage du dispositif. La dernière était la session de test.

### 5.4.5.1 Principaux résultats

Les observations lors des sessions d'apprentissage mettent bien en évidence une amélioration des performances du sujet. Leur temps de parcours diminue et ils sortent de moins en moins du parcours. En revanche, aucun effet lié à la simulation de la voie Magno ON lors des phases d'apprentissage n'a pu être mis en évidence.

Dans la session de test, on observe bien un effet comportemental induit par le dispositif. Dans le cas du groupe contrôle, cet effet s'observe sur la différence relative de temps de parcours. Les sujets sont plus rapides en condition normale qu'en condition inversée. Cette différence est toutefois très petite, de l'ordre de 4 %. Par ailleurs, on n'observe pas d'effet significatif de l'inversion de l'image sur le nombre de contacts.

Dans le groupe "Magno", c'est sur la différence relative de nombre de contacts que l'on observe un effet significatif. Cette différence relative est par ailleurs assez importante, de l'ordre de 35%. L'effet induit par le dispositif sur le temps de parcours, de l'ordre de 5%, n'est quant à lui pas significatif, mais une tendance se dessine (le test de Student donne un  $p$  inférieur à 0.1).

Le fait que la simulation de la voie Magno augmente l'effet induit par le dispositif reste toutefois à confirmer. En effet, la différence entre l'effet induit chez les sujets du groupe Magno et chez ceux du groupe contrôle n'est pas significative.

### 5.4.5.2 Relation entre apprentissage et performances au test

Pour savoir si la mesure de l'effet induit par le dispositif dans la phase de test fait écho ou non à l'amélioration des performances des sujets dans la phase d'apprentissage, nous avons tracé la mesure de différence relative de nombre de contacts des sujets dans la phase de test, en fonction de la pente de la courbe d'apprentissage pour le nombre de contacts, plus exactement l'opposé de cette pente, de manière à ce qu'un fort apprentissage conduise à des fortes valeurs du taux d'apprentissage, et en fonction du groupe (figure 5.17)

Dans le cas du groupe Magno, la relation entre le taux d'apprentissage et la mesure de RNCD est quasi linéaire. Une régression linéaire sur les données indique par ailleurs que la corrélation est

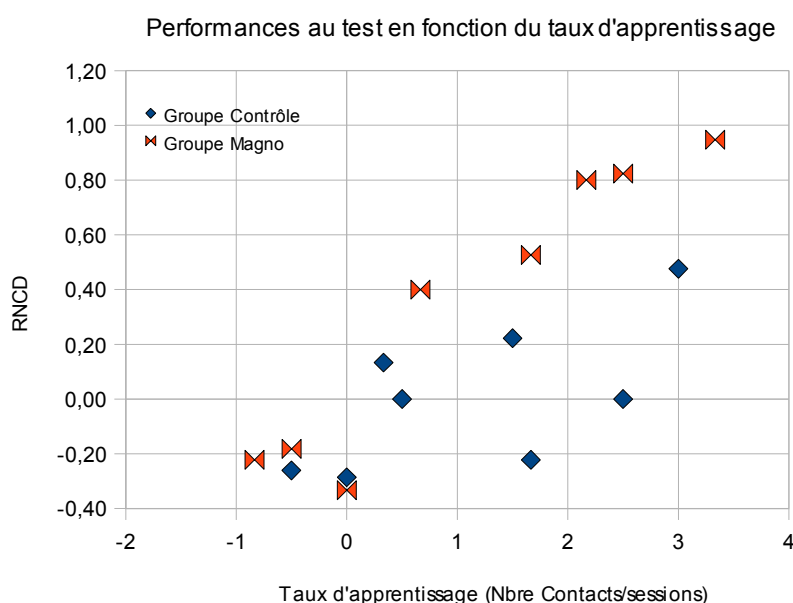


Figure 5.17 : Corrélation entre le taux d'apprentissage mesuré sur le nombre de contacts, et la mesure de la différence relative de contacts lors du test.

significative ( $p < 0.01$ ). Cette forte corrélation entre les phases d'apprentissage et la phase de test indique que l'apprentissage des sujets est bel et bien lié à l'utilisation du dispositif, et que la mesure de différence relative de performance est un indicateur valable de l'efficacité du dispositif.

Dans le cas du groupe contrôle, la pente est beaucoup plus faible et la corrélation n'est pas significative ( $p > 0.05$ ), ce qui suggère un effet très limité du dispositif.

#### 5.4.5.3 Analyse du comportement des sujets

Nous allons maintenant proposer une analyse du comportement des sujets par la méthode des répondants. La discussion qui va suivre est issue de deux sources :

- De l'observation directe du comportement des sujets lors des sessions d'apprentissage et de test
- D'un questionnaire que nous avons distribué aux sujets à l'issue de l'expérience et dans lequel nous leur demandions d'explicitier leurs stratégies, et d'expliquer, de leur point de vue, le fonctionnement du dispositif. Ces questionnaires sont reproduits en annexes.

Dans un premier temps, on identifie deux types de sujets : les sujets "passifs" et les sujets "actifs". Les sujets passifs (1, 2, 9, 12 et 15) se laissent porter par le son. Leur tête est en générale immobile, fixée droit devant ou à quelques mètres au sol devant eux.

Lorsqu'on les interroge sur leurs stratégies, les sujets passifs répondent en général qu'ils ont tenté d'identifier un son particulier correspondant au chemin à suivre comme par exemple le sujet 2 : "J'ai essayé de repérer un son spécifique... et j'ai essayé de le suivre", ou le sujet 9 : "J'ai essayé de trouver une relation auditive, percevoir les variations du son afin d'en établir des codes directionnels".

On retrouve aussi cette idée de manière implicite dans la formulation d'hypothèses concernant le fonctionnement du dispositif (Sujet 15 : "Quant il y a un obstacle, les sons sont plus fort, ils

ronronnent en quelque sorte quand vous allez dans la bonne direction"), ou dans la formulation de remarques concernant l'amélioration du dispositif (Sujet 15 : "J'aurais souhaité être avertie des obstacles par un son plus identifiable que ceux entendus. J'aurais voulu que le dispositif m'indique si le chemin allait plus à droite ou plus à gauche", Sujet 1 qui s'était renseigné sur le dispositif après l'expérience : "Je suis étonné que ça ne fonctionne pas en détection d'objet").

L'analyse en première personne du comportement des sujets passifs pourrait donc être formulée ainsi : "un dispositif sonore m'indique le chemin à suivre et les obstacles à éviter. Je dois apprendre quel son correspond à quelle information". Nous l'appellerons "stratégie du guidage sonore".

Les sujets passifs, enfin, ont en général des performances médiocres. Cela peut se comprendre dans la mesure où ils n'ont pas compris le principe de fonctionnement du dispositif (sans parler du codage bien sûr).

Chez les autres sujets, l'activité se traduit par deux choses : des mouvements de la tête fréquents, parfois continus, de type balayage, et chez certains sujets, des pauses dans la mobilité pour "écouter l'endroit où l'on est", comme le précise le sujet 10.

Chez les sujets actifs, le rapport au dispositif est pensé de manière interactive, comme une adaptation mutuelle : le sujet 5 par exemple explique qu'il "aurait aimé apprivoiser [...] le dispositif en [se] baladant n'importe où".

La plupart des sujets actifs ont par ailleurs bien compris que le son était lié directement à l'image de la caméra, et que le dispositif n'était pas un dispositif de guidage : "Je dirais qu'il y a comme caractéristiques principales la luminosité [...] et la couleur" (sujet 7), "J'imagine que le son se différencie en fonction de la matière filmée par la caméra (végétaux, béton, ciel...)" (Sujet 16). L'idée d'un dispositif de guidage est toutefois présente chez certains sujets actifs (par exemple chez le sujet 3 : "le son de la bonne direction semblait être à une gamme de fréquence donnée, sans pour autant être agréable"). Chez certains sujets actifs enfin, la compréhension du dispositif mêle les deux idées : "Il me semble que chaque canal RVB est traduit en fréquence [...] et l'effet stéréo est dû à une détection de mouvement général dans l'image" (sujet 5). Ce dernier sujet propose par ailleurs dans les commentaires de "Gommer ce tissage fréquentiel un peu perturbant" mais il rajoute ensuite "Mais l'avantage d'un tel son brut est qu'avec l'expérience il peut sûrement être interprété dans n'importe quel contexte".

Dans le groupe des actifs, on trouve essentiellement deux types de stratégies. La première consiste à associer des sons à des endroits particuliers et à les mémoriser : "A chaque fois j'essayais de repérer un son particulier et l'endroit où je devais me situer par rapport à lui (soleil, bâtiments...)" (sujet 16), "Au début je tournais la tête dans tous les sens pour "regarder" autour de moi mais ensuite j'ai plutôt gardé la même position[...] pour pouvoir reconnaître l'endroit.[...].

L'analyse en première personne de la stratégie de ces sujets serait donc : "Le dispositif que j'utilise traduit l'image de la caméra en son. Ainsi, à chaque image correspond un son particulier. Je dois donc apprendre à reconnaître des images particulières qui jalonnent mon parcours pour m'en servir de points de repère". Nous appellerons cette méthode "la stratégie du cliché".

Cette stratégie présente cependant des inconvénients majeurs : "les sons étaient différents selon les endroits mais pas assez pour pouvoir se repérer" (S16), "à la fin j'arrivais à reconnaître certains passages assez bien. Par contre si j'avais le malheur de me planter je me retrouvais vite perdu dans un environnement auditif inconnu et ininterprétable" (S5).

La majeure partie des sujets actifs ont opté pour une autre stratégie. Elle consiste à faire des mouvements permanents avec la tête de manière à "accrocher" de l'information, c'est à dire à

identifier une variation reproductible et caractéristique dans le signal. La variation la plus fréquemment utilisée était celle qu'il y avait entre le chemin, et l'herbe qui borde le chemin.

L'évolution décrite par le sujet 8 est assez caractéristique : "Lors des premiers parcours, j'ai tenté de regarder face à moi, comme si je marchais normalement à vue. Bien qu'il y avait des différences dans les sons, il était impossible d'accrocher un repère intéressant à suivre, "l'image" étant trop floue pour distinguer à longue distance. Puis j'ai tenté la distance intermédiaire. Il était alors possible de repérer les textures unies sans trop de difficulté. Au fur et à mesure, je me suis rendu compte qu'il valait mieux regarder très près de soi (1-2 mètres devant ses pieds)[...].J'ai [alors] distingué l'herbe [de l'accotement] du chemin [ce qui m'a permis de] suivre ce dernier en balayant de la tête ce qui se situait devant moi sur un angle d'environ 45°".

Par ces mouvements permanents, le sujet suit en fait la bordure du chemin. On retrouve cette même stratégie chez plusieurs sujets, la plupart ayant de bons résultats (S7, S10, S11, S13).

Le sujet 13 par ailleurs utilisait le même principe pour se diriger vers un pylône situé au niveau de l'arrivée. Nous reviendrons sur cela par la suite.

L'analyse en première personne de cette dernière catégorie de sujet pourrait donc être formulée ainsi : "Le son produit par le dispositif change lorsque je fait des mouvements. Après quelque temps, je m'aperçois que dans certaines situations, ces changements sont réguliers c'est à dire que je peux les reproduire cycliquement par un mouvement particulier. Je me sers alors de ces régularités pour me guider". Nous appellerons cette méthode "la stratégie de l'accroche"

On notera ici que le processus identifié est très similaire à celui des sujets de Charles Lenay qui, pour identifier une lettre sur une table graphique, effectuent avec le stylet des oscillations gauche-droite perpendiculaires au contour suivi. Dans le cas présent, le sujets effectuent aussi des oscillations gauche droite autour de la frontière chemin/herbe, et se déplacent parallèlement à cette frontière.

On remarquera enfin qu'en général, les sujets combinent cette stratégie de l'accroche avec une stratégie du cliché, cette dernière servant surtout à l'orientation. Cela est explicité par le sujet 8 qui ajoute à la fin du paragraphe reproduit plus haut : "De temps à autre, je levais la tête afin de repérer les arbres (ou ce que je pensais être des arbres) afin d'avoir une idée de l'avancement sur le chemin". Nous reviendrons sur cette question par la suite

#### 5.4.5.4 A la recherche de répondants en troisième personne

Nous allons maintenant, à partir de notre propre connaissance et de notre propre expérience du dispositif, tâcher d'identifier des "répondants" en troisième personne aux stratégies identifiées dans le paragraphe précédent. Dans un premier temps, nous allons considérer les capacités du dispositif en l'absence d'un traitement préalable de l'image.

##### **Sans traitement de l'image :**

La stratégie du guidage tout d'abord. Nous pensons que cette stratégie est en premier lieu liée à un pré-supposé sur le fonctionnement du dispositif. Ce pré-supposé initial est ensuite corroboré par certaines particularités du parcours. En particulier, le chemin sur lequel le sujet doit évoluer est constitué de dalles claires homogènes et sans obstacles alors que les pourtour est plus sombre et présente de nombreux éléments (banc, arbres, bâtiments). Il s'ensuit naturellement que lorsque je pointe la caméra vers le chemin, le son est plus fort et présente toutes les fréquences du spectre. Il constitue un bruit blanc caractéristique qui peut être interprété comme un signal de guidage (le

"bon" son). Avec cette stratégie, il est bel et bien possible de reconnaître le chemin. Cette stratégie n'est en revanche d'aucune efficacité aux endroits du parcours où il y a plusieurs chemins possibles, ou lorsque l'on se situe au milieu d'une "place".

Pour ce qui concerne la stratégie du cliché ensuite, nous pensons qu'elle est également liée à un présupposé, d'ordre épistémologique cette fois : elle correspond à l'idée que le cerveau va enregistrer passivement un son, le mémoriser, et le reconnaître au moment voulu. Si l'on considère cette stratégie au pied de la lettre, c'est à dire en considérant qu'il s'agit de reconnaître une image par son codage sonore, elle s'avère totalement inefficace pour réaliser la tâche. Les sons produits par le dispositif sont en effet très sensibles à une toute petite variation de la pose, de sorte qu'il s'avère impossible d'identifier auditivement une scène visuelle donnée d'un parcours sur l'autre. Rappelons par ailleurs à ce sujet que la capacité de reconnaissance d'objets y compris en environnement simple et contrôlé est très limitée si le sujet ne peut pas bouger la caméra (Bach-y-Rita, 1972). Que dire donc d'une scène visuelle entière et en extérieur. En revanche, il est effectivement possible de distinguer des zones en fonction de leur luminance. Les plages d'herbes sont en général plus sombres que le chemin. Si je fixe la caméra sur le chemin, le son est donc plus fort que si je la fixe sur l'herbe. Les capacités de discrimination à l'aide de cette méthode sont donc limitées à des différences de luminance. Par ailleurs, le chemin à l'ombre peut avoir une intensité plus faible que l'herbe au soleil. Nous verrons dans la suite du propos que le traitement de la voie Magno permet d'aller au delà de la simple discrimination de plages de luminance.

Voyons maintenant ce qu'il en est de la stratégie de l'accroche. Lorsque l'on effectue des mouvements d'oscillation perpendiculaires au bord du chemin, on obtient effectivement un phénomène reproductible : une alternance entre un son fort (le chemin de couleur clair) et un son faible (l'herbe de couleur sombre). A partir de ce seul indice, il est effectivement possible de longer le bord du chemin. Dans l'ombre néanmoins, la différence entre les deux amplitudes devient imperceptible. Par ailleurs, cette stratégie pose problème lorsque les variations de luminosité sont liées à d'autres phénomènes que l'alternance herbe/chemin. Par ailleurs, une ombre sur le chemin peut être prise à tort pour un tournant, ce qui explique les pauses marquées parfois par les sujets devant ces dernières.

En dehors du cas particulier de l'herbe et du chemin, la stratégie de l'accroche permet d'identifier à distance certaines structures de grande taille comme par exemple les pylônes ou certains immeubles présentant une structure en bandes claires et foncées. L'important est que la structure forme une bande ou une alternance de bandes horizontales lumineuses. Ce type de structure active tous les champs récepteurs dans une bande donnée de l'image, c'est à dire toutes les fréquences du grave à l'aigu. Cela produit un bruit blanc caractéristique dont la source apparente correspond à la direction du pylône. Dans le cas d'un bâtiment, ce bruit blanc se déplace avec le mouvement de la tête, apparaissant dans une oreille pour disparaître dans l'autre et faire place à un nouveau bruit blanc. Le son correspond en quelque sorte à celui d'un livre qu'on feuillette.

### **Avec traitement de l'image :**

Voyons maintenant comment le dispositif se comporte dans ces mêmes cas de figure lorsque le traitement de l'image est activé, c'est à dire que l'on simule le comportement de la voie Magno.

Le premier constat est que la stratégie du cliché et du guidage ne sont plus du tout efficaces. En effet, si je ne fais pas de mouvements particuliers, le filtrage passe-bande simulant les cellules Magno atténue considérablement les différences de luminosité entre le chemin et l'herbe (fig. 5.18 et 5.19). Par ailleurs, si je me mets à marcher en maintenant la caméra devant moi, comme dans la stratégie du guidage, le signal devient plus fort mais il est difficile à utiliser en tant que tel car il ne



Figure 5.18 : image d'origine issue de la caméra



Figure 5.19 : image prétraitée dans le cas où le sujet est parfaitement immobile



Figure 5.20 : arrêt sur image du chemin filmé en faisant un balayage gauche droite avec la caméra. Le sens de la flèche indique le sens de mouvement de la caméra.

présente pas de régularités. J'ai donc abordé rapidement les stratégies de type "accroche".

Si je dirige la caméra à quelques pas devant moi et que j'effectue un mouvement de balayage gauche-droite, j'obtiens cette fois un phénomène reproductible intéressant. Lorsque je fait un balayage vers la gauche, j'obtiens un signal caractéristique dans l'oreille gauche. A l'inverse, lorsque je fais un balayage vers la droite, j'obtiens un signal caractéristique dans l'oreille droite. Je m'aperçois rapidement que ce signal correspond aux bordures droites et gauches du chemin.

Pour étudier ce phénomène, nous avons réalisé un film du parcours en effectuant un balayage gauche-droite avec la caméra. Nous avons ensuite appliqué à ce film le même filtrage que celui implémenté dans le dispositif. La figure 5.20 présente deux arrêts sur image de ce film, lors d'un balayage gauche/droite et lors d'un balayage droite/gauche.

Le son caractéristique correspond à la bande blanche à gauche ou à droite. L'apparition de cette bande est directement liée au prétraitement de l'image. Nous avons mentionné dans le chapitre 2 que les cellules de la voie Magnocellulaire répondaient de manière transitoire à des variations du signal. Dans le cas d'un mouvement, c'est au niveau des frontières entre des zones différentes de l'image

que ces changements seront les plus importants. La frontière entre le chemin et l'herbe est une de ces frontières, c'est pourquoi elle apparaît lors d'un mouvement de balayage.

Le lecteur remarquera néanmoins que les bords gauche et droits du parcours n'apparaissent pas simultanément. Cela est lié à la polarité des cellules Magno, ici de type ON.

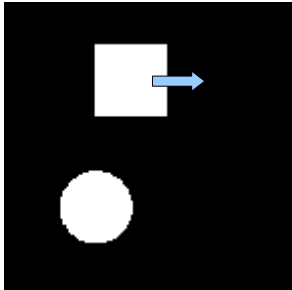


Figure 5.21 : Dessin animé constitué d'un carré blanc en mouvement et d'un cercle blanc fixe

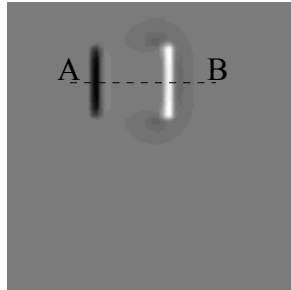


Figure 5.22 : Arrêt sur image de la simulation du dessin animé dans la voie Magno (recombinée).

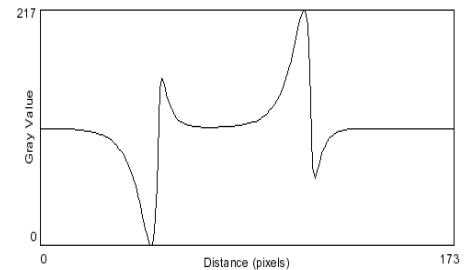


Figure 5.23 : Niveaux de gris de l'image Magno recombinaison sur le profil A-B

Pour illustrer le propos, nous avons réalisé un film constitué d'un carré et d'un rond. Le rond est fixe. Le carré est animé d'un mouvement de la gauche vers la droite (figure 5.21). Nous simulons ensuite le prétraitement associé à la voie Magnocellulaire (ON et OFF recombinaison) sur ce film. La figure 5.22 montre un arrêt sur image de ce film prétraité. Le gris correspond au niveau 0. Les valeurs au-dessus de ce niveau de gris correspondent à l'activation des cellules ON. Celles en dessous à l'activation des cellules OFF. La courbe de la figure 5.23 présente la valeur des niveaux de gris mesurée sur le profil A-B.

Comme on peut le constater sur la figure 5.22, le cercle, immobile, disparaît totalement. De la même manière, la partie interne du carré ainsi que ses bords supérieurs et inférieurs (qui sont colinéaires à la direction de mouvement) disparaissent. Seuls subsistent les zones de changement, en l'occurrence les bords gauche et droits du carré, qui sont orthogonaux au déplacement. Les valeurs sur le bord droit sont supérieures au niveau moyen et correspondent à une activation des cellules ON. Le bord gauche présente lui des valeurs inférieures au niveau moyen, il correspond donc à l'activation de l'autre voie : la voie OFF. Si l'on ne considère que la voie ON, on obtiendra donc :

- Dans le cas d'un mouvement de la caméra vers la gauche, du signal sur le bord droit du carré seulement
- Dans le cas d'un mouvement de la caméra vers la droite, du signal sur le bord gauche du carré seulement

Une fois cette régularité repérée, il est relativement facile de se maintenir au centre du chemin en effectuant des mouvements de balayage avec la tête. C'est entre autre cette méthode qu'applique notamment le sujet 11, qui est l'un de nos meilleurs sujets (et le seul à avoir réussi un sans faute lors de l'apprentissage). Une technique similaire permettrait au sujet 13 de repérer les pylônes à la fin du parcours. Ces pylônes sont visibles sur la figure 5.20, ils encadrent le porche. Un balayage gauche droit les fait très clairement apparaître.

Cette méthode de balayage gauche-droite n'est toutefois pas la seule stratégie ouverte par le dispositif. Une autre stratégie permet de détecter ce que les sujets nomment des "textures". Les



sujets y font souvent référence, soit explicitement (S8), soit implicitement (le sujet 13 explique qu'il peut différencier les "surfaces lisses", le sujet 16 pense que le son dépend de "la matière filmée par la caméra").

Ces textures n'apparaissent pas lorsque l'on se concentre pour maintenir la caméra immobile. Au contraire, elles apparaissent lorsque l'on effectue des petits mouvements avec la tête. Ces mouvements ne sont pas nécessairement dans une direction donnée, des mouvements circulaires ou même désordonnés conviennent également. Ces mouvements peuvent par ailleurs être uniquement liés au déplacement.

Le mieux pour comprendre cette impression de "texture" est de se référer au sens tactile. Si on me demande d'éprouver la texture d'une matière, ma stratégie va être de déplacer mon doigt sur cette surface. Supposons que la matière soit lisse, ce déplacement provoquera peu de changements sensoriels. Si en revanche la surface est rugueuse, je sentirai de nombreux changements. Si maintenant on me demande simplement de poser mon doigt sur la surface, je suis incapable de décrire sa texture. C'est dans ce sens qu'il y a une analogie entre une texture tactile et la "texture" d'une zone testée avec le dispositif.

Les "textures" que l'on identifie correspondent donc à des variations du signal sonore liées à des mouvements de faible amplitude. Contrairement au cas du balayage gauche-droite, l'identification de ces textures ne nécessite pas de faire un mouvement stéréotypique. Par ailleurs, ce qui est identifiable n'est pas un son particulier mais un pattern sonore, une caractéristique dans les variations du son. Cette texture change effectivement en fonction de ce que l'on regarde. En particulier, le béton du chemin, la pelouse, et les arbres alentours présentent des "textures" différentes.

La caractéristique du béton est précisément qu'il ne génère pas de variations sonores lorsqu'on fait des petits mouvements de caméra : d'où cette impression d'une "surface lisse" (S13). En revanche, la pelouse, regardée d'en haut, et les arbres regardés de loin, produisent des variations caractéristiques et semblent avoir une "texture".

Les arbres vus de profil semblent particulièrement "rugueux". Des petits mouvements de la tête provoquent immédiatement l'activation de nombreuses sources sonores réparties sans règle identifiable dans tout le champ perceptif. Une comparaison possible serait ce type particulier de feux d'artifice qui produisent de très nombreuses étincelles à des positions imprévisibles. Quand on arrête le mouvement, les "étincelles" s'évanouissent.

L'herbe, enfin, semble moins rugueuse que les arbres. Les variations sonores sont moins denses et moins fréquentes. On pourrait comparer la structure des variations sonores induites par le mouvement de la caméra aux bulles qui apparaissent à la surface de l'eau lorsqu'elle commence à bouillir (le sujet 13 appelait ce phénomène les "bloupbloup"). *A posteriori*, les sujets ont souvent cru que ce son était lié aux marguerites dans l'herbe, ce qui donne une idée de la densité des sources sonores impliquées : plutôt clairsemées. Encore une fois, ces patterns sonores disparaissent si l'on se concentre pour immobiliser son "regard". Le sujet 13 en a pâti à ses dépens. Dans la dernière séance en effet, nous informons les sujets qu'il s'agit de la séance de test et qu'ils doivent être particulièrement vigilants. A l'issue du test, le sujet 13 s'est plaint de n'avoir pas réussi, justement cette fois, à repérer la pelouse. Lorsque j'ai par la suite expliqué au sujet d'où provenait ces "textures" et qu'elle étaient liées au mouvement de la caméra, ce dernier a alors compris pourquoi il n'avait pas réussi à les retrouver : parce qu'il se concentrait pour maintenir la position de la caméra fixe, pensant que c'était la bonne méthode pour repérer ces fameuses "marguerites".

Voyons maintenant si l'on peut identifier à quoi correspondent ces "textures" en terme de signal.

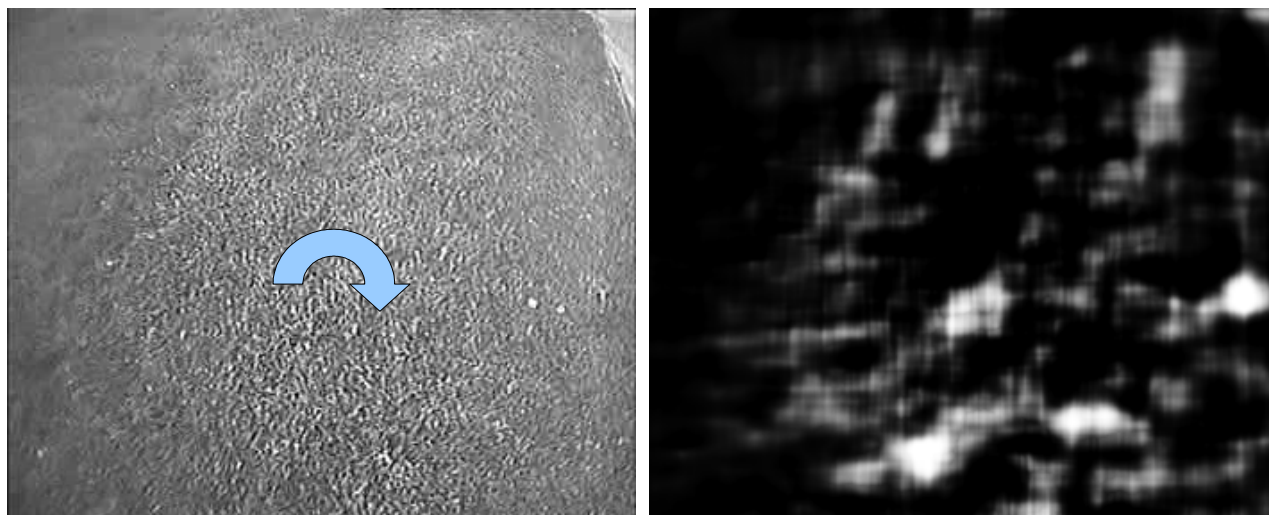


Figure 5.24 : la pelouse "vue" à l'aide de légers mouvements du dispositif (à droite). Ce type de tache apparaissent puis disparaissent rapidement à des positions variées dans l'image. A gauche, l'image initial.

Pour cela nous avons filmé la pelouse du terrain en faisant des légers mouvements avec la caméra. Nous avons ensuite simulé le prétraitement effectué par le dispositif. La figure 5.24 présente le résultat obtenu. Lorsque l'on fait des petits mouvements avec la caméra, des taches éparses apparaissent puis disparaissent dans l'image. Leur position est imprévisible, en particulier, il ne s'agit pas de taches qui se déplacent. On retrouve ce même type de tache au niveau de la pelouse sur la figure 5.24. Ce sont ces taches qui activent en général une seule source sonore chacune. Ces sources sonores ne sont pas nécessairement contiguës. Elles s'activent puis se désactivent rapidement, comme des bulles qui apparaissent puis éclatent à la surface de l'eau bouillante.

Remarquons pour terminer que l'identification de "textures" n'est pas l'apanage des sujets du groupe Magno. C'est notamment le sujet 8, du groupe contrôle, qui fait explicitement référence à cette notion. Il indique cependant "repérer les textures unies sans trop de difficultés". Or le critère pour repérer une surface unie est que le signal sonore ne change pas, ce qui est vrai aussi bien avec que sans pré-traitement de l'image. On peut néanmoins supposer que si le sujet fait référence à une notion de surface lisse, c'est par opposition à d'autres surfaces qui ne le sont pas, mais qu'il a toutefois du mal à identifier. Dans son questionnaire il fait en particulier référence à des arbres. Or les arbres possèdent naturellement une structure similaire aux structure produites ici par la pelouse : le soleil au travers des feuilles forment des taches lumineuses éparses qui ressemblent à celles de la figure 5.24. Lorsque l'on regarde un arbre avec le dispositif sans traitement de l'image, on obtient effectivement une sensation similaire de texture. Son origine est toutefois très différente. En particulier, dans le cas d'un arbre et du dispositif du groupe contrôle, les taches lumineuses restent stables et constantes.

#### 5.4.5.5 Différences entre l'utilisation du dispositif avec et sans pré-traitement de l'image

A l'issue de cette analyse en première et troisième personne du fonctionnement du dispositif, il convient de faire le point sur les différences liées au traitement du signal dans la manière de manipuler le dispositif.

Il s'avère tout d'abord que la simulation de la voie Magno pousse naturellement à l'utilisation d'une stratégie d'accroche. Ceci s'explique bien sûr par le comportement transitoire de la voie

Magno : sans mouvement du sujet, le signal est faible. Les autres stratégies s'avèrent totalement inefficaces dans ce cas.

Dans le cadre des stratégies d'accroche, identifie-t-on des différences dans le comportement des sujets? A première vue, non. En effet, on retrouve dans les deux groupes le même type de balayage gauche-droite continu, les sujets des deux groupes sont capables de séparer le chemin de la pelouse alentour, et on retrouve dans les deux groupes des références à la notion de texture.

Cette similitude en surface n'est cependant qu'apparente. L'analyse par la méthode des répondants permet en effet de constater des différences importantes dans les réactions du dispositif en réponse aux mouvements des sujets. Par exemple dans le cas du balayage gauche-droite de la tête, le dispositif du groupe contrôle produit une alternance signal fort/signal faible. Pour le même type de mouvement, le dispositif du groupe Magno produit, lui, un signal à la frontière des deux zones, qui n'apparaît par ailleurs que dans une des deux directions du mouvement. De plus, si dans le groupe contrôle on est capable d'identifier des zones composées de tâches lumineuses (comme par exemple un arbre) en effectuant des petits mouvements, ces petits mouvements dans le groupe Magno produisent des tâches que l'on est capable d'identifier à la manière d'une texture. Ainsi, si dans les deux cas les sujets se basent sur les variations du signal pour réaliser la tâche, il subsiste une différence de taille entre le fonctionnement du dispositif du groupe contrôle et dans celui du groupe Magno. Dans le premier cas le signal est reçu par le sujet, qui réalise des actions pour en tirer une information. Dans le second cas, les variations du signal sont produites par l'action du sujet, qui en tire alors une information. Le premier cas correspond à un processus sensorimoteur pur. Le second correspond à mon sens au concept d'énaction défini par Varela.

La question de savoir quels sont les rapports entre les deux approches, et en particulier si elles sont totalement équivalentes ou non est aujourd'hui en débat. Je n'ai pour ma part ni le bagage ni le recul nécessaire pour aborder cette question aujourd'hui. Voyons maintenant ce que l'on peut conclure de cette analyse quant au rôle joué par le traitement de l'image dans le dispositif.

#### 5.4.5.6 Rôle de la simulation de la voie Magno ON dans le dispositif

Il est au final difficile de déduire à partir de cette analyse du comportement des sujets le rôle joué par la simulation de la voie Magno ON dans la prise en main et l'utilisation du dispositif. Il s'avère qu'avec ou sans filtrage, les sujets sont capables d'identifier des variations liées à des actions qui s'avèrent utiles pour réaliser la tâche.

Dans un cadre passif, on considère usuellement que le filtrage associé à la voie Magno permet de "sélectionner" l'information importante pour l'analyse du mouvement et la mobilité. Comme nous l'avons vu au chapitre 2, la voie Magno est en effet un filtre passe bande spatio-temporel qui véhicule essentiellement les variations du signal au niveau de la rétine.

Cette idée n'est pas incompatible avec une approche sensorimotrice. En effet, cette approche repose sur l'idée que la perception émane du lien entre les variations du signal et des variations motrices. A ce titre, le fait de sélectionner uniquement les variations dans le signal fait tout à fait sens.

Cet aspect est d'ailleurs effectivement appréciable lorsque l'on utilise le dispositif : le confort d'écoute est bien meilleur lorsque le prétraitement est activé que lorsqu'il ne l'est pas. Ce prétraitement permet en effet d'éliminer une bonne partie du "bruit de fond" du dispositif, c'est à dire l'ensemble des signaux qui ne varient pas de manière sensible lorsque l'on effectue des mouvements avec le dispositif. Il correspond en particulier aux grandes plages lumineuses unies (ciel, réflexions sur les bâtiments, etc...). Ces plages disparaissent lorsqu'on active le prétraitement,

et il devient plus aisé de distinguer les variations du signal.

Toutefois, à l'issu de cette analyse des différentes stratégies des sujets, ce point de vue me paraît très partiel. Il ne reflète pas en particulier la richesse qui émane des nouvelles formes d'interactions ouverte par ce "traitement", qui s'apparente d'ailleurs plus à une forme de "comportement" qu'à une notion de filtrage, au sens où le dispositif réagit d'une manière très particulière.

## 5.4.6 Conclusion

Dans cette expérience, nous avons testé l'effet de l'implémentation d'une simulation de la voie Magnocellulaire de type ON du primate dans le dispositif de substitution sensorielle TheVIBE sur le comportement de sujets voyants ayant les yeux bandés dans une tâche de mobilité. La tâche de mobilité consistait en un parcours sans obstacle dans un environnement extérieur.

Après une phase d'apprentissage, nous testons l'effet d'une inversion de l'image de la caméra sur les performances du sujet. Cette inversion a bien un effet significatif chez les sujets du groupe test (simulation de la voie Magno) et du groupe contrôle. Dans le groupe test, la taille de cet effet est plus importante que dans le groupe contrôle. Par ailleurs la corrélation entre le taux d'apprentissage des sujets est significatif dans le groupe test mais ne l'est pas dans le groupe contrôle. L'effet du filtrage n'est cependant pas significatif.

Dans le but de mieux comprendre le rôle du traitement implémenté dans le dispositif, nous nous sommes penché sur le comportement des sujets et plus particulièrement sur leurs stratégies. Il s'avère que, d'un point de vue extérieur, ces stratégies sont relativement similaires : elle consistent à effectuer des mouvements avec la caméra et à identifier des régularités dans les variations du signal sonore. Dans ce cadre, l'un des intérêts du prétraitement de l'image choisi est qu'il sélectionne précisément les variations dans l'image et filtre ce qui reste constant. Au regard des stratégies employées par les sujets, cette analyse semble cependant partielle. Le prétraitement effectué permet en effet au sujet d'expérimenter des formes d'interaction nouvelles avec le dispositif : à la notion de filtre spatio-temporel on préférera donc la notion de comportement spatio-temporel.

## Références

- Auvray, M.; Hanneton, S.; Lenay, C. & O'Regan, K. (2005), 'There is something out there: distal attribution in sensory substitution, twenty years later', *Journal of Integrative Neuroscience* **4**(4), 505.
- Cha, K.; Horch, K. W. & Normann, R. A. (1992), 'Mobility performance with a pixelized vision system.', *Vision Res* **32**(7), 1367--1372.
- Dagnelie, G.; Keane, P.; Narla, V.; Yang, L.; Weiland, J. & Humayun, M. (2007), 'Real and virtual mobility performance in simulated prosthetic vision.', *J Neural Eng* **4**(1), S92--101.
- Dowling, J. (2005), 'Artificial human vision', *Expert Rev Med Devices* **2**(1), 73--85.
- Gibson, J. (1986), *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates.
- Jansson, G. (1986), *Development of electronic aids for the visually impaired...*, Dordrecht: Martinus Nijhoff, chapter Development and evaluation of mobility aids for the visuallyhandicapped.

Kallie, C. S.; Schrater, P. R. & Legge, G. E. (2007), 'Variability in stepping direction explains the veering behavior of blind walkers.', *J Exp Psychol Hum Percept Perform* **33**(1), 183--200.

Leat, S. J. & Lovie-Kitchin, J. E. (2006), 'Measuring mobility performance: experience gained in designing a mobility course.', *Clin Exp Optom* **89**(4), 215--228.

Sander, M.; Bournot, M.; Lelièvre, F. & Tallec, D. A. (2005), 'Les personnes en situation de handicap visuel. Apports de l'enquête Handicaps-Incapacités-Dépendances', Rapport d'enquête du Ministère Français de la Santé et des Solidarités.

Shapiro, S. & Wilk, M. (1965), 'An analysis of variance test for normality (complete samples)', *Biometrika* **52**(3,4), 591-611.

Shingledecker, C. A. & Foulke, E. (1978), 'A human factors approach to the assessment of the mobility of blind pedestrians.', *Hum Factors* **20**(3), 273--286.

West, S. K.; Rubin, G. S.; Broman, A. T.; Muñoz, B.; Bandeen-Roche, K. & Turano, K. (2002), 'How does visual impairment affect performance on tasks of everyday life? The SEE Project. Salisbury Eye Evaluation.', *Arch Ophthalmol* **120**(6), 774--780.

## Chapitre 6 : Conclusions et Perspectives

L'objectif de cette thèse était d'implémenter une simulation de la chaîne rétino-corticale du système visuel humain dans une prothèse visuelle afin d'améliorer les performances du dispositif et d'en étudier le rôle dans l'utilisation du dispositif. Le cadre de cette étude était celui de la théorie des contingences sensori-motrices. Résumons maintenant les contributions de ce travail.

### 6.1 Conclusions

#### 6.1.1 Contributions techniques

En pratique, nous avons tout d'abord mis en place une simulation logicielle du traitement de l'image dans la chaîne rétino-corticale du primate reproduisant : la densité spatialement variante des photorécepteurs à la surface de la rétine, l'adaptation de la réponse des cellules de la rétine en fonction des conditions locales de luminance et de contraste, et enfin l'aiguillage du signal visuel en fonction de ses composantes fréquentielles spatiotemporelles au niveau du nerf optique, formant respectivement les voies Parvo et Magnocellulaires (chapitre 2). Cette simulation repose sur un modèle développé au laboratoire depuis 1994. L'apport essentiel de ce travail a été de compiler les travaux réalisés sur différents maillons de cette chaîne et de les assembler en une seule simulation. Cette simulation est par ailleurs entièrement paramétrable de manière à être adaptée à différentes applications et en particulier à la vision prosthétisée. Cette simulation représente enfin une étude de faisabilité : dans le cas de la vision prosthétisée, il s'agissait en effet de montrer que l'on peut simuler ce traitement à une cadence vidéo standard, ce qui s'avère être le cas.

Cette simulation a donné lieu à plusieurs publications et présentations (Durette *et al.* 2005 , 2006a, 2006b, 2007, 2009 ; Chaix de Lavarène *et al.* 2007 ; Hérault et Durette, 2007 ; Benoit *et al.* 2007, 2009)

En pratique nous avons également participé à l'amélioration d'un dispositif de substitution sensorielle récent dénommé TheVIBE.

En premier lieu, nous avons proposé des méthodes innovantes permettant de mieux maîtriser le processus d'échantillonnage dans le dispositif (chapitre 3) grâce aux méthodes d'auto-organisation de Kohonen. Cette méthode permet d'améliorer la couverture de l'image avec un nombre limité d'échantillons.

Nous avons également proposé une analyse des enjeux du codage sonore dans le dispositif et notamment du choix des "échelles de codage sonore", en lien avec la théorie des contingences sensori-motrices. Nous avons en particulier montré comment la théorie de l'information, en lien avec la notion de "canal sensoriel" en psychoacoustique permet de déterminer des ordres de grandeurs sur le nombre de points envisageable lorsque l'on utilise une interface sonore. Nous avons pour cela proposé une méthode de calcul de la capacité d'un canal sensoriel reposant sur les capacités de discrimination du sujet, et non plus sur sa capacité de catégorisation (chapitre 3).

Finalement, nous avons développé une nouvelle version du logiciel dans lequel l'image peut-être prétraitée avant d'être échantillonnée.

En pratique enfin, nous avons mis en place des protocoles d'apprentissage et d'évaluation du dispositif. Ces protocoles sont graduels et tiennent compte du processus d'immersion d'un sujet dans

un dispositif de substitution. Ils permettent de travailler sur les premières étapes d'immersion qui concernent l'apprentissage du couplage sensori-moteur induit par le dispositif.

### 6.1.2 Contributions sur la question du handicap

Les premiers protocoles concernent la capacité de diriger la caméra du dispositif vers une cible puis la capacité d'aller chercher cette cible dans une salle (Chapitre 4). Nous avons testé des sujets voyants ayant les yeux bandés dans ce type de tâche. Nous avons par ailleurs testé l'effet d'un échantillonnage spatialement variant de l'image de la caméra reproduisant l'échantillonnage de l'image rétinienne par la distribution des photorécepteurs sur la rétine. Dans le cas de la tâche de pointage de la cible avec la caméra, les performances des sujets dans le groupe spatialement variant se sont avérées nettement supérieures à celles du groupe contrôle. Dans la tâche de préhension en revanche, nous n'avons pas observé de différences significatives entre les performances des deux groupes.

Dans la suite nous nous sommes concentrés sur la question de la mobilité, dont l'enjeu pour la communauté des non-voyants est de premier ordre (Chapitre 5). Nous avons choisi pour cela un contexte et une tâche écologique : le sujet devait apprendre à réaliser un parcours dans un environnement réel (un garage dans l'expérience préliminaire, les allées d'une cours dans l'expérience princeps).

L'une des conclusions importantes de cette étude est que le dispositif de substitution a bien un effet sur le comportement des sujets dans une tâche de mobilité. Ceci permet de relancer l'idée qu'un dispositif de substitution visuelle (à ne pas confondre avec un dispositif de substitution en général, la canne blanche peut par exemple être considérée comme un dispositif de substitution et a largement fait ses preuves en matière d'aide à la mobilité) est bien un candidat pour une aide à la mobilité, ce qui apparaissait comme très improbable jusqu'alors (voir par exemple Kaczmarek, 1995)

Autre observation importante dans le cadre de cette étude : il semble que le traitement du signal soit un préliminaire indispensable à l'obtention d'un effet comportemental du dispositif, au moins en tout cas dans un environnement extérieur. Nous avons en effet testé l'effet de la simulation de la voie "Magno", considérée en générale comme la voie dédiée aux informations de contexte et de mouvement, dans le dispositif pour une tâche de mobilité en extérieur. En environnement extérieur, nous obtenons un effet significatif du dispositif sur le comportement des sujets du groupe bénéficiant de la simulation de la voie Magno, alors que cet effet est beaucoup plus limité dans le cas du groupe contrôle. C'est donc potentiellement parce qu'aucun prétraitement n'était effectué sur l'image que les dispositifs de substitution visuelle ont jusqu'à aujourd'hui été considérés comme inefficaces dans le cadre de la mobilité.

Ces travaux ont également donné lieu à plusieurs articles et présentations (Durette *et al.*, 2007, 2008)

### 6.1.3 Contributions sur le plan de la recherche

Sur le plan de la recherche, nous avons en premier lieu mis en évidence en pratique les enjeux et les conséquences du choix d'une approche théorique de la vision (en l'occurrence passive versus active) sur la conception d'une prothèse.

Au niveau du dispositif, nous avons montré en particulier les différences impliquées par le fait de considérer la variation du signal et non sa valeur nominale comme élément signifiant dans le signal

(en lien avec une variation motrice bien sûr). Nous avons montré les conséquences de cette considération sur le choix des échelles de codage sonore, sur la mesure de la capacité d'un canal sensoriel et sur le nombre de points admissible dans le cas de notre interface. Nous avons enfin proposé une analyse mathématique du rôle de la notion d'ordre dans les échelles choisies, et du sens du choix de fonctions de conversion visuo-sonore monotones, toujours dans un cadre sensori-moteur.

L'autre contribution de ce travail de thèse consiste en des propositions concernant le rôle du traitement du signal dans les dispositifs de perception dans un cadre sensori-moteur. Si certaines sont issues de considérations théoriques, comme par exemple l'intérêt d'un échantillonnage logarithmique et d'une projection polaire, d'autres sont d'origine empirique : elles proviennent de l'analyse du comportement des sujets, de leur commentaires vis à vis de l'utilisation du dispositif, et de notre propre expérience avec le dispositif. C'est le cas du rôle du "filtrage" spatio-temporel opéré dans la rétine, auquel nous préférons le qualificatif de "comportement" spatio-temporel. Ce comportement spatio-temporel joue vraisemblablement un rôle crucial, notamment dans la perception des textures. Il constitue de ce point de vue un enrichissement du couplage sensori-moteur (Chapitre 5). A ce titre, cette étude peut être considérée comme un préliminaire à la détermination de nouvelles contingences sensori-motrices liées au traitement du signal dans le dispositif, et à l'identification de nouveaux répondants liés à ces même contingences.

## 6.2 Perspectives

Les perspectives de ce travail de thèse sont nombreuses, tant sur le plan de la prise en charge du handicap que sur le plan de la recherche fondamentale.

### 6.2.1 Sur le plan du handicap

Sur le plan du handicap, les perspectives à court terme concernent le logiciel TheVIBE. La version mise au point au laboratoire va permettre tout d'abord de tester l'intérêt des autres prétraitements proposés dans cette thèse, en l'occurrence les adaptations locales dans la rétine, et la transformation rétinocorticale Log-Polaire.

Par ailleurs, la méthode d'évaluation mise en place dans le cadre de la mobilité à la particularité d'être sensible ce qui permet d'obtenir une mesure de l'efficacité du couplage sensorimoteur induit en peu de temps. Elle pourra donc être utilisée pour améliorer le réglage des paramètres du traitement du signal, dans une boucle de conception rapide.

Enfin, l'une des perspectives essentielles aujourd'hui est de tester le dispositif chez des sujets non-voyants de naissance. En collaboration avec l'association Valentin Haüy, nous sommes allés rencontrer des sujets non-voyant à qui nous avons fait essayer le dispositif. A cette occasion, nous avons constaté une grande curiosité pour le dispositif, en particulier chez les sujets non-voyant de naissance, et ce après quelques minutes d'utilisation seulement. Cet intérêt ne portait pas seulement sur les aspects pratiques du dispositif mais également sur l'interaction qu'il ouvrait avec l'environnement. A l'intérêt pratique du dispositif s'ajoute donc un intérêt d'un point de vue ludique.

L'utilisation du sens auditif chez des sujets non-voyants reste toutefois une étape, car l'ouïe est un sens d'une importance capitale chez l'aveugle et nous pensons qu'un dispositif auditif risque de parasiter cette entrée. Dans le cadre du projet SSSHV, en collaboration avec Vincent Fristot de l'équipe architecture du DIS, nous avons montré la faisabilité d'une matrice de stimulation tactile pilotée en tension. Nous avons pour cela construit un premier dispositif de stimulation et mesuré



les premières courbes psychométriques sur des sujets (fig. 1 et 2). Un prototype d'interface comprenant 64 électrodes est aujourd'hui en phase de tests. Cette matrice se placera sur l'avant-bras.

Par ailleurs, le travail sur la substitution sensorielle présente de nombreuses questions communes avec les systèmes invasifs. Il constitue donc une bonne plateforme de tests préliminaires.



Figure 1 : test de faisabilité du stimulateur tactile

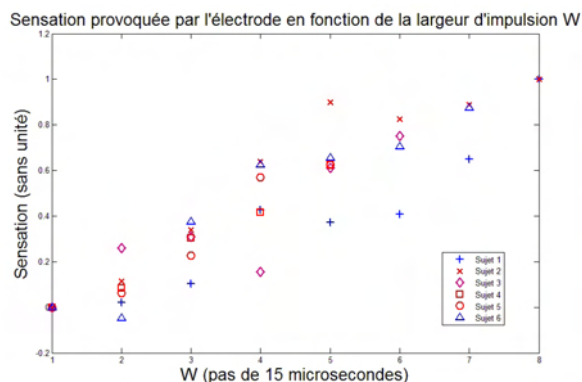


Figure 2 : fonctions psychométriques de la sensation tactile en fonction des paramètres du signal de stimulation

## 6.2.2 Sur le plan de la recherche

Les dispositifs de substitution constituent des outils très intéressants sur le plan de la recherche car ils permettent de s'interroger sur le fondement des processus perceptifs et en particulier sur leur genèse. Associés à la méthode des répondants, ils permettent de déceler leurs véritables principes de fonctionnement, en mettant en évidence des comportements effectivement mis en oeuvre par les utilisateurs pour constituer un environnement perceptif à l'aide du dispositif.

Sur le plan de la recherche, la question qui se pose est bien sûr de savoir dans quelle mesure on peut faire un parallèle entre le fonctionnement de ces dispositifs et les systèmes visuels naturels. Nous avons vu par exemple dans le chapitre 5 que la plupart des sujets effectuaient des mouvements continus et réguliers pour se servir du dispositif. On peut donc se demander si ce type de mouvements existe dans les systèmes naturels.

Chez l'Homme, on a identifié depuis longtemps des mouvements permanents de l'oeil. Ces mouvements sont connus sous le nom de micromouvements (pour une revue récente, voir par exemple Martinez-Conde *et al.*, 2004). Ils sont rapides et de très faible amplitude, de l'ordre d'un déplacement de l'image sur la rétine de quelques photorécepteurs. Par ailleurs, ils ont lieu y compris lorsque le sujet fixe son regard. On sait enfin depuis longtemps que si on asservit une image par rapport à ces micromouvements de l'oeil, le sujet ne la perçoit plus.

Ces micromouvements, en particulier les microtremblements, ont pendant longtemps été considérés comme des mouvements parasites, aléatoires, liés à l'activité spontanée des muscles oculaires. Des résultats récents permettent aujourd'hui de remettre cette idée en question. Une étude de Spaschus *et al.* (1999) montre par exemple que les microtremblements de l'oeil gauche et de l'oeil droit sont corrélés. Par ailleurs, Rucci *et al.* (2007) montrent que les micromouvements ont un rôle dans la capacité du système visuel à détecter des "gratings" (réseau périodique) dans du bruit. Il

sera à ce titre intéressant de comparer ces mouvements continus et de faible amplitude observés dans les systèmes perceptifs naturels avec les mouvements de caméra effectués par un sujet équipé d'un dispositif de substitution visuelle.

## Références

Benoit, A.; Caplier, A.; Durette, B. & Hérault, J. (2009), 'Using human visual system modeling for bio-inspired low level image processing', *Computer vision and Image understanding*.

Benoit, A.; Caplier, A.; Durette, B. & Hérault, J. (2007), 'From bio-inspired vision toolbox to realtime computer vision algorithms', European Retina Meeting 2007, Frankfurt, Germany.

Chaix de Lavarène, B.; Alleysson, D.; Durette, B. & Hérault, J. (2007), 'Efficient demosaicing through recursive filtering' IEEE International Conference on Image Processing'.

Durette, B.; Alleysson, D. & Hérault, J. (2009), 'Simulation of the retina : a tool for visual prostheses', *International Journal of Image and Graphics*, sous presse

Durette, B.; Benoit, A.; de Lavarène, B. C.; Alleysson, D. & Hérault, J. (2007), 'A biomimetic real time retinal simulation software with non separable spatiotemporal processing, chromaticity and non linear local adaptivity', European Retina Meeting, Frankfurt, Germany.

Durette, B.; Benoit, A.; de Lavarène, B. C.; Alleysson, D. & Hérault, J. (2006), 'Real-time bio-inspired retina simulation software' Advances in Neural Information Processing Systems'.

Durette, B.; Corvino, R.; Mancini, S.; Alleysson, D. & Hérault, J. (2006), 'Model of the human retina for sensory substitution and retinal implants' Euroconférence "Sensory perception: basic mechanisms and human diseases", Institut Pasteur.

Durette, B.; Gamond, L.; Hanneton, S.; Alleysson, D. & Hérault, J. (2007), 'Biomimetic space-variant sampling in a vision prosthesis improves the user's skill in a localization task' 5th Conference and Workshop on Assistive Technology for People with Vision and Hearing Impairments', Granada, Spain.

Durette, B.; Louveton, N.; Alleysson, D. & Hérault, J. (2008). (2008), 'Visuo-auditory sensory substitution for mobility assistance : testing TheVIBE' Workshop on Computer Vision for Applications for the Visually Impaired, 10th European Conference on Computer Vision (ECCV'2008)', Marseille, France.

Hérault, J. and Durette, B. (2007), 'Modeling Visual Perception for Image Processing' WANN 2007', 662-675.

Kaczmarek, K. & y Rita, P. B. (1995), *Advanced Interface Design and Virtual Environments*, Oxford University Press, chapter Tactile Displays, pp. 349-414.

Martinez-Conde S., Macknik, S.L., an Hubel D.H..(2004) The rôle of fixational eye movements in visual perception. *Nature Neuroscience* 5(3), p 229-240

Rucci, M. Lovin, R., Poletti M. and Santini, F. (2007) Miniature eye movements enhance fine spatial detail. *Nature*, 447, p851-855

Spaschus A., Marsden J., Halliday, D.M., Rosenberg, J.M. And Brown, P. (1999) The origin of ocular microtremor in man. *Exp Brain Res*, 126, p 556-562.

# ANNEXES



## Annexe du chapitre 2

### 1 Projection pseudo-logarithmique

On modélise l'échantillonnage logarithmique par une loi pseudo logarithmique de la forme :

$$\rho' = \rho_{\text{lim}}' \frac{\rho}{\rho + \rho_0}$$

qui lie  $\rho'$ , la distance du photorécepteur sur la rétine, à  $\rho$  la distance du point qu'il code sur l'image ;  $\rho_{\text{lim}}$  est une constante et  $\rho_0$  un paramètre qui influe sur la force de la compression log.

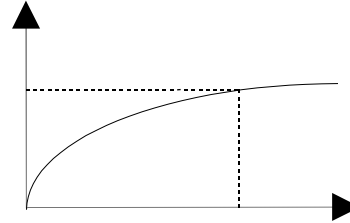


Figure 1 : Fonction d'échantillonnage  $\rho' = f(\rho)$

De l'équation  $\rho' = \rho_{\text{lim}}' \frac{\rho}{\rho + \rho_0}$  on déduit  $\begin{pmatrix} x \\ y \end{pmatrix} = \frac{\rho_0}{\rho_{\text{lim}}' - \rho'} \begin{pmatrix} x' \\ y' \end{pmatrix}$

Ainsi, pour chaque pixel de l'image échantillonnée de coordonnée  $(x', y')$ , on va chercher dans l'image le pixel  $(x, y)$  qui lui correspond.

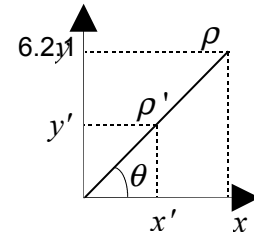


Figure 2 : notation pour l'echt log

On choisi, de plus, de réduire l'image d'un facteur "k". Soit W la largeur de l'image, le point de coordonné  $(W/2, 0)$  qui correspond au bord de l'image, doit devenir le point  $(W/(2*k), 0)$  dans l'image échantillonnée. Cette contrainte permet de déduire  $\rho_{\text{lim}}'$  :

$$\frac{W}{2} = \frac{\rho_0 \cdot \frac{W}{2k}}{\rho_{\text{lim}}' - \frac{W}{2k}} \Rightarrow \rho_{\text{lim}}' = \frac{1}{k} \left( \frac{W}{2} + \rho_0 \right)$$

La fonction "ech\_log (X,nfact,p0)" implémente cet échantillonnage logarithmique, avec nfact le paramètre de réduction de l'image (d'un facteur  $2^{\text{nfact}}$ ), et p0 de l'ordre de la taille de l'image d'arrivée. Ajoutons que les zones où l'échantillonnage sortira de l'image de départ (zone blanche sur l'image échantillonnée ci-dessous) sont remplies par "ech\_log" qui y place les pixels de l'image d'origine par une transformation linéaire simple (homothétie).

La figure ci dessus montre le résultat de l'échantillonnage de l'image "Lena" réduite d'un facteur 2 , avec et sans "effet de bord" pour 2 valeurs de p0 (256 et 128).

Figure 3 : Echantillonnage de "Lena" , réduction d'un facteur 2. 1) image initiale ; 2) et 3) eht log pour  $\rho_0=256$  avec et sans remplissage de la bordure ; 4) et 5) idem pour  $\rho_0=128$



Cependant, dans cette version de l'échantillonnage, le paramètre  $\rho_0$  n'a pas le même effet suivant la taille de l'image, c'est pourquoi nous proposons une version adimensionnée de l'implémentation de l'échantillonnage spatialement variant.

## 2 Version adimensionnée

On peut écrire l'équation de transformation de la manière suivante :

$$\xi = \alpha_{\lim} \frac{x}{x + \alpha_0}, \text{ avec } (\rho', \rho, \rho_{\lim}, \rho_0) = \frac{W}{2}(\xi, x, \alpha_{\lim}, \alpha_0), W \text{ étant la taille de l'image}$$

1<sup>ère</sup> condition limite :

1- Soit  $k$  le facteur de réduction de l'image, on veut que le point au bord de l'image initiale soit projeté au bord de l'image réduite. Cette condition ne peut être réalisée partout car l'image est de forme rectangulaire. On prendra un des cotés.

$$\frac{1}{k} = \alpha_{\lim} \frac{1}{1 + \alpha_0} \text{ donne } \alpha_{\lim} = \frac{1 + \alpha_0}{k}$$

## 3 Densité d'échantillonnage

Dans l'espace projeté, on écrit le nombre d'échantillon par unité de volume :

$$dn = \lambda dV_{\xi} \text{ avec } dV_{\xi} = \xi d\xi d\theta$$

$$\xi = \alpha_{\lim} \frac{x}{x + \alpha_0} \text{ donne } d\xi = \frac{\alpha_{\lim} \alpha_0}{(x + \alpha_0)^2} dx$$

$$dn = \lambda dV_{\xi} = \lambda \alpha_{\lim} \frac{x}{x + \alpha_0} \frac{\alpha_{\lim} \alpha_0}{(x + \alpha_0)^2} dx d\theta$$

$$dn = \lambda \frac{\alpha_{\lim}^2 \alpha_0}{(x + \alpha_0)^3} dV_x$$

D'où la densité d'échantillonnage sur l'image d'origine :

$$f(x) = \lambda \frac{\alpha_{\lim}^2 \alpha_0}{(x + \alpha_0)^3}$$

2<sup>ème</sup> condition aux limites :

Par hypothèse, la densité d'échantillonnage dans l'espace projetée est uniforme et égale à  $\lambda$ .

On en déduit le rapport du taux d'échantillonnage dans l'image initiale sur le taux d'échantillonnage dans l'image projetée:

$$R(x) = \frac{\alpha_{\lim}^2 \alpha_0}{(x + \alpha_0)^3}$$

$$\text{En } 0, \text{ ce rapport vaut : } R(0) = \frac{\alpha_{\lim}^2}{\alpha_0^2}$$

$$\text{De plus, on a } \alpha_{\lim} = \frac{1 + \alpha_0}{k}, \text{ d'où } R(0) = \frac{(1 + \alpha_0)^2}{k^2 \alpha_0^2}$$

on note  $R(0) = Ro$ , l'équation précédente s'écrit :

$$(k^2 Ro - 1) \alpha_0^2 - 2 \alpha_0^2 - 1 = 0$$

Elle admet deux solutions réelles dont on ne gardera que la solution positive :

$$\alpha_0 = \frac{(1 + k\sqrt{Ro})}{(k^2 Ro - 1)} \text{ qui donne la valeur du paramètre } \alpha_0 \text{ en fonction du rapport des taux d'échantillonnage en } 0.$$

$$\text{Par exemple pour un rapport de } 1 \text{ (1 pixel donne 1 pixel), on obtient } \alpha_0 = \frac{1}{k - 1}$$

Dans la fonction ech\_log\_adim, k et Ro sont des parametres librement modifiable



## 4 Filtrage progressif avant échantillonnage

L'idée est de filtrer l'image avec un filtre passe bas spatialement variant avant l'échantillonnage de manière à respecter localement le théorème de Shannon. On utilisera pour ça les fonctions de densité d'échantillonnage préalablement calculées : la fréquence de coupure du filtre étant égale à la moitié de la densité d'échantillonnage locale.

### 4.1 Filtrage progressif avec un filtre RII du premier ordre

On souhaite utiliser un filtre du même type que ceux décrit pour la simulation d'une couche de corps cellulaires dans la rétine (cf. rapport de stage durette\_2005) :

$$G(z_{k1}, z_{k2}) = \left( \frac{1-a}{1+a} \right)^2 \cdot \left( \frac{1}{1-az_{k1}^{-1}} + \frac{1}{1-az_{k1}} - 1 \right) \cdot \left( \frac{1}{1-az_{k2}^{-1}} + \frac{1}{1-az_{k2}} - 1 \right)$$

Au niveau de l'implémentation à ce niveau, il faut penser à distribuer le facteur  $\left( \frac{1-a}{1+a} \right)^2$  sur toutes les termes de l'équation, sous peine d'avoir des problèmes de gain.

#### 4.1.1 Calcul du paramètre a

Le paramètre a de filtrage va cette fois dépendre de la position dans l'image ( en l'occurrence de l'excentricité dans l'image). Nous devons donc calculer a(e) tel que la fréquence de coupure fc(e) de notre filtre soit égale à f<sub>ech</sub>(e)/2.

L'hypothèse est que cette fréquence de coupure varie lentement par rapport à l'échantillonnage.

#### 4.1.2 Paramètre a en fonction de fc

Soit gc le gain du filtre pour f=fc, on cherche à résoudre :

$$G(gc)=fc$$

En première approximation, on travaillera en 1 dimension. L'équation devient, en fréquence réduite :

$$\left( \frac{1-a}{1+a} \right) \frac{1-a^2}{1+a^2-2a\cos(2\pi f_c)} = g_c$$

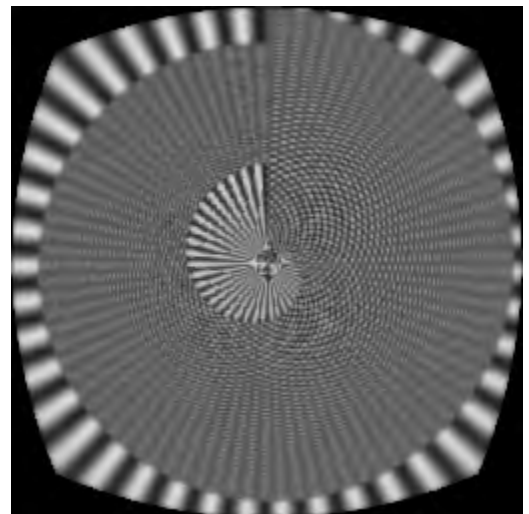
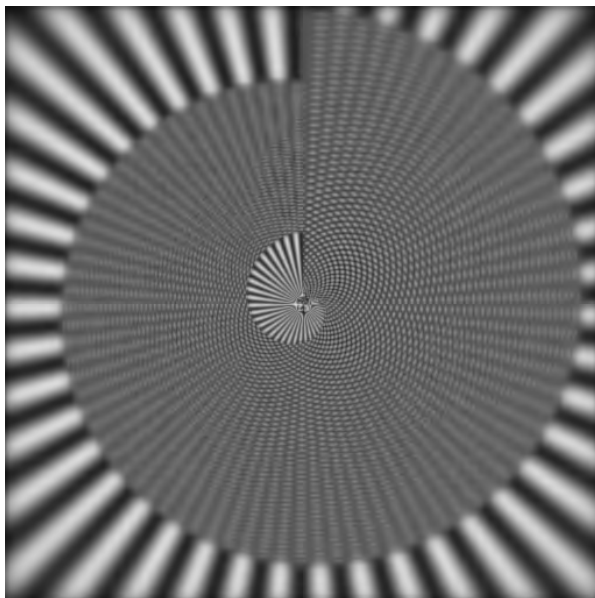
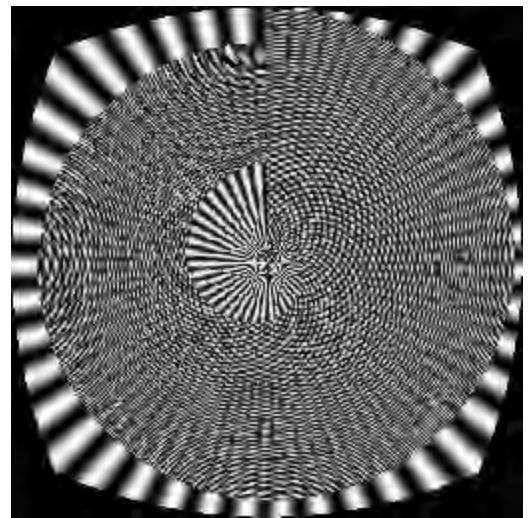
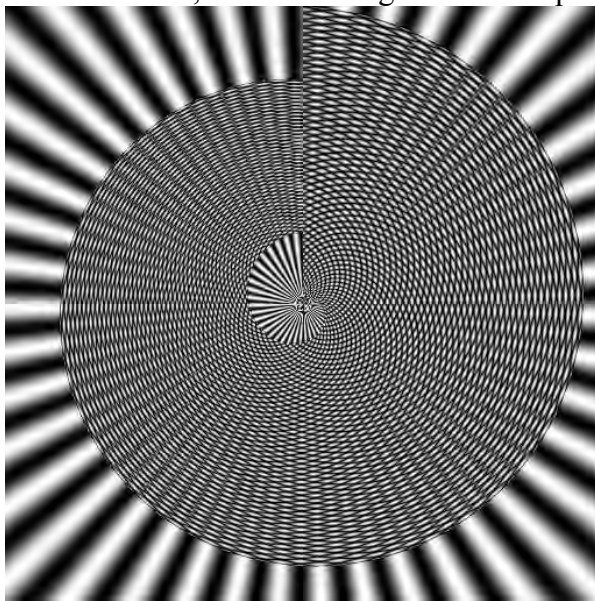
On choisira la solution telle que a<1 (merci,maxcima)

$$a = \frac{\sqrt{(\cos(2\pi f_c)^2 - 1)g_c^2 + (2 - 2\cos(2\pi f_c))g_c + \cos(2\pi f_c)g_c - 1}}{g_c - 1}$$

La fonction `init_a_new_new` implémente le calcul du paramètre  $a$  à partir d'une matrice de la distance des différents pixels dans l'image, normalisée (i.e. entre 0 et 1), pour une densité d'échantillonnage correspondant à la projection proposée au paragraphe 2. Il est possible de régler le paramètre `gc` dans le code.

#### 4.1.3 Résultats

Le script `script_fil_ech_2` permet de comparer les images obtenues avec et sans filtrage préalable. La comparaison peut être faite directement entre les images projetées, ou après rétablissement d'une grille d'échantillonnage régulière à l'aide des fonctions `inv_ech_log.m`, `inv_ech_log_rand.m` ou encore `inv_ech_log_interp.m`. Cette dernière fonction en particulier réalise une interpolation par spline des valeurs des pixels dans l'image. On peut également s'intéresser au spectre de l'image. A titre indicatif, certaines images sont ici reproduites.



Sous échantillonnage d'une mire. A gauche mire d'origine (en haut image originale, en bas image ayant subi un filtrage spatialement variant). A droite mire sous-échantillonnée.

## Annexe du chapitre 3

### Exemple d'un fichier "properties.ini" d'option de TheVIBE :

```
[video]
device = 0                # 0 - one camera
freq = 30.                # framerate
width = 320               # image width
height = 240              # image height
config = 1                # 0 - no camera config; 1 - camera config
invert = 0                # 0 - no image black / white inversion; 1 - inversion

[sound]
chunkSize = 1024          # size of the sound chunk
freq = 22050              # sound sampling frequency

[save]
yes = 0                   # save the results
period = 1                # save period
fileName = savetest.txt   # save file name

[retina]
threshold = 32            # [0 - 255] - under tis threshold no sound at all
fileName = ret200_4.txt   # retina definition file

[filter]
on = 1                    # 0, 1
compression = 0           # 0, 1
k = 6.0                   # [0.001 - 20]
tau = 2.0                 # [0 - 20]
beta = 0.2                # [0 - 1]
V0 = 32.0                 # [0-255]

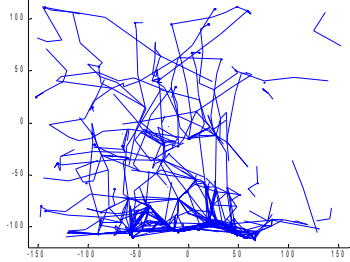
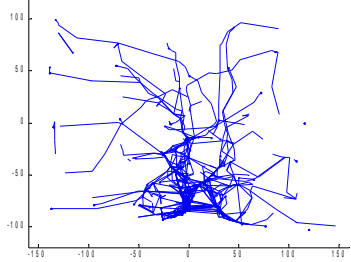
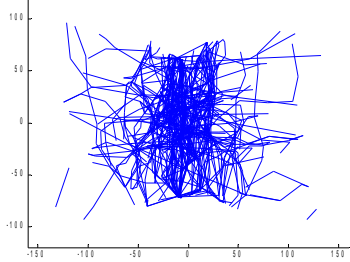
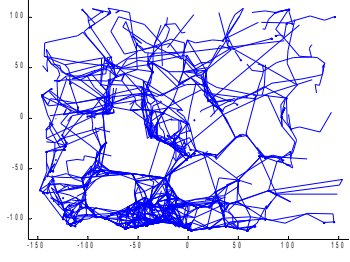
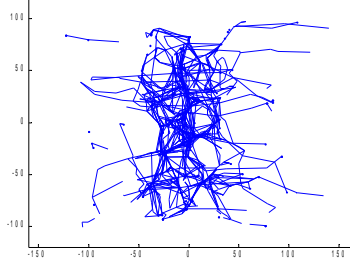
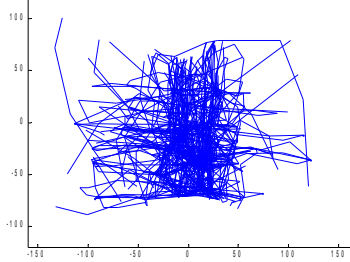
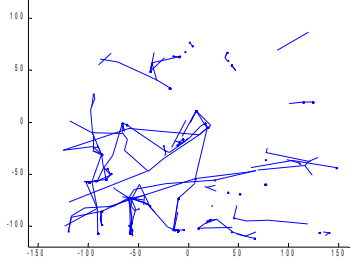
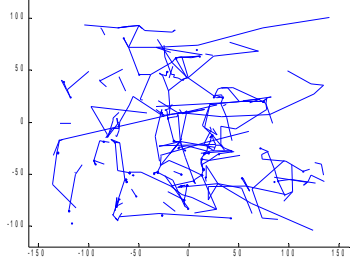
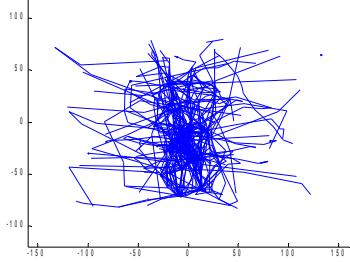
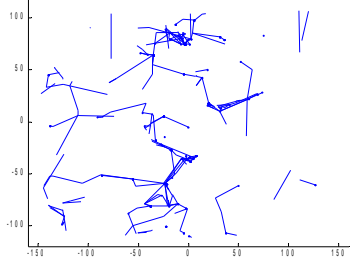
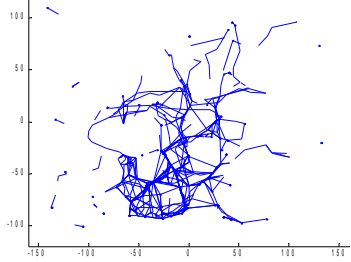
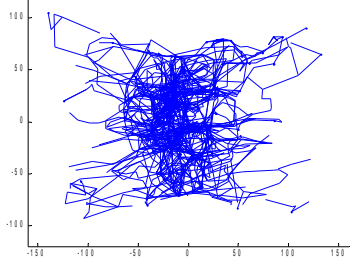
[mode]
expe = 1                  # 0, 1 // on off

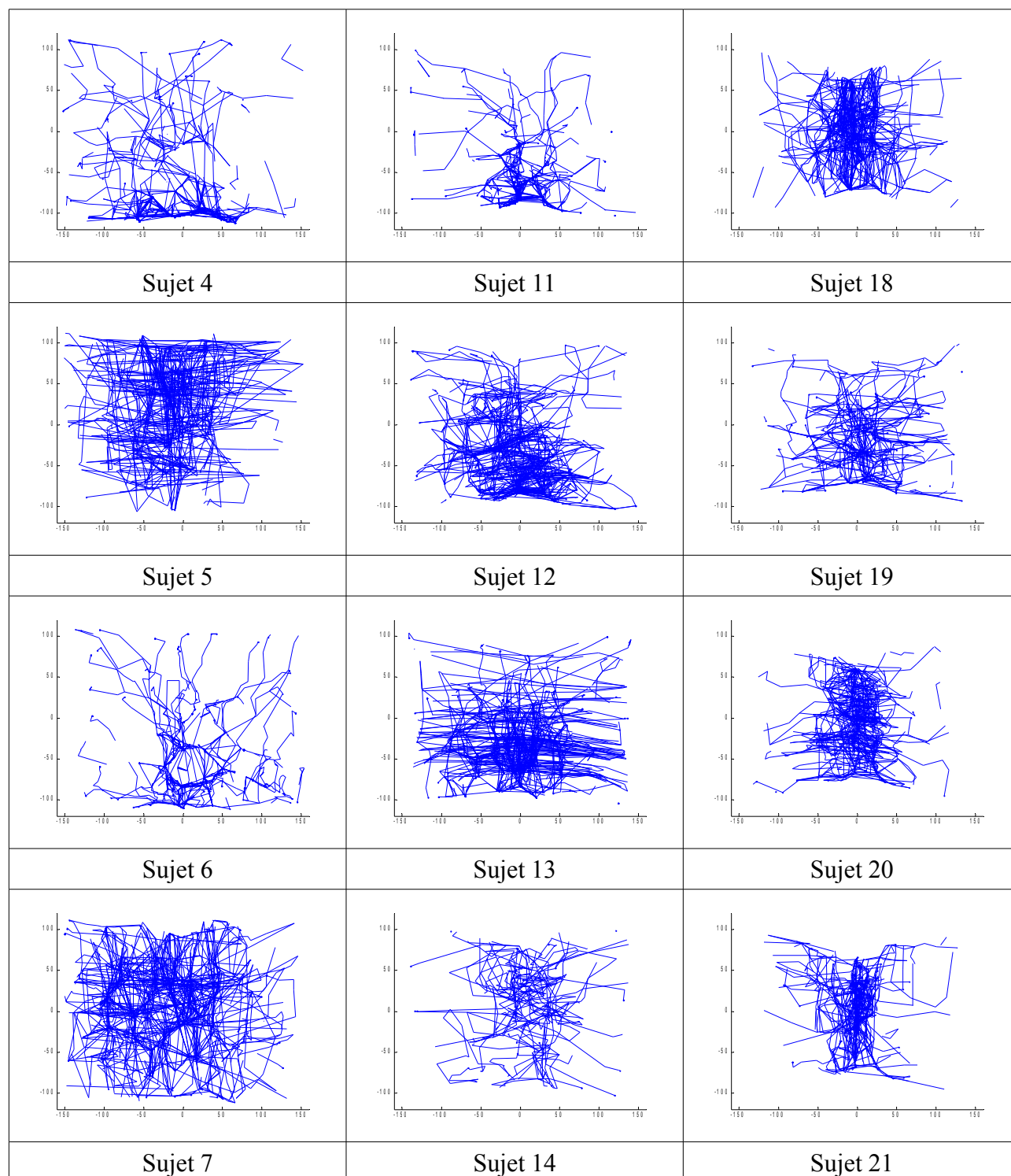
[image]
mode = 1                  # 0 = RGB ; 1 = YUY2
```



## Annexe du chapitre 4

### Expérience 1 : Déplacement de la cible dans le champ de la caméra lors de l'expérience de saccades visuelle

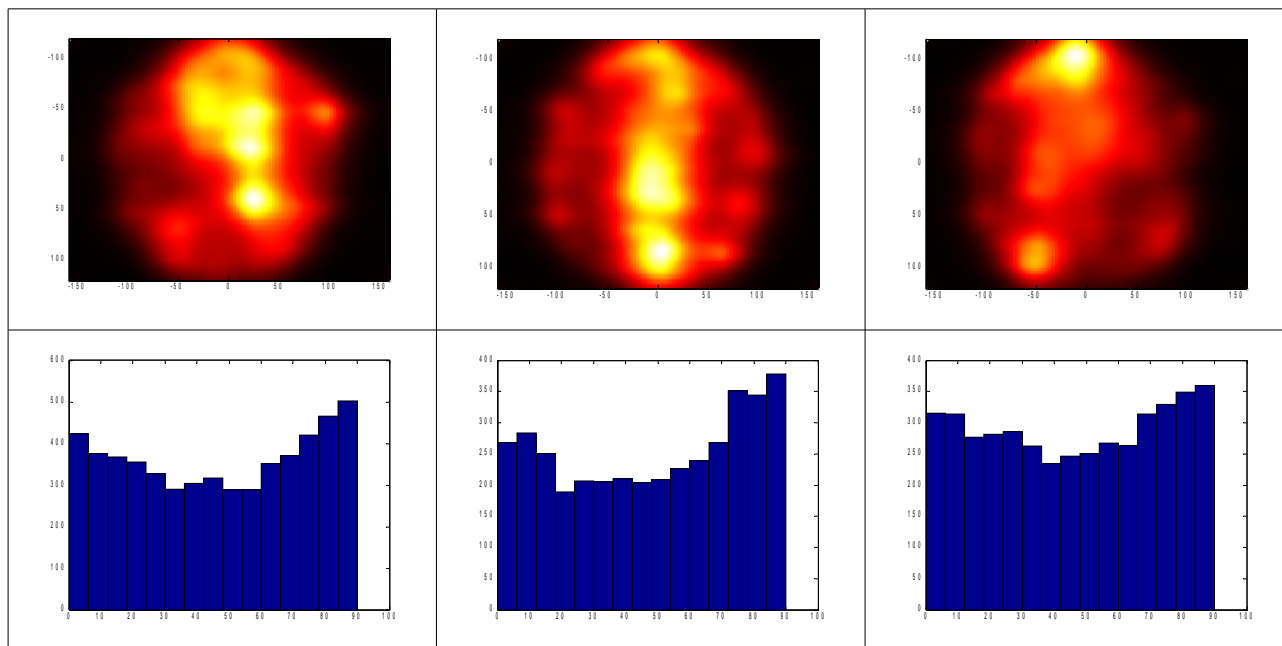
Groupe Unif (dans l'ordre des sujets)	Groupe Log2 (dans l'ordre des sujets)	Groupe Log 4 (dans l'ordre des sujets)
		
Sujet 1	Sujet 8	Sujet 15
		
Sujet 2	Sujet 9	Sujet 16
		
Sujet 3	Sujet 10	Sujet 17
		



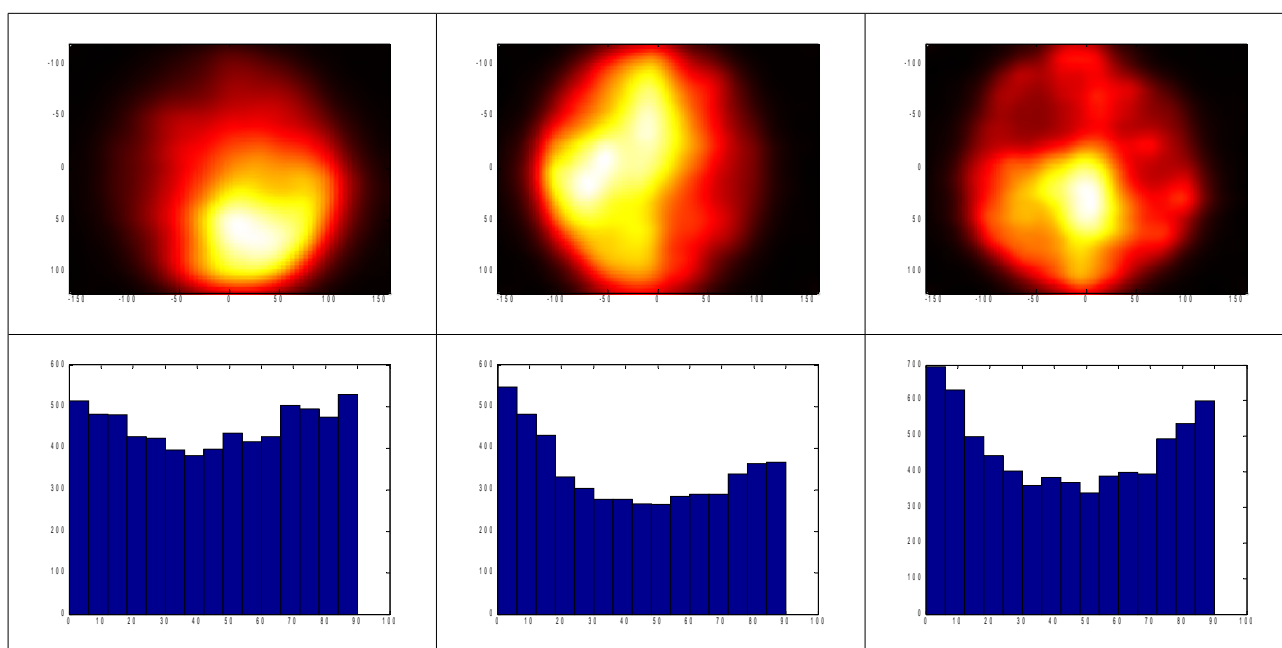
## Expérience 2 : Densité des positions de la cible, et histogramme de l'orientation des mouvements de la cible dans le champ de la caméra

### Groupe Uniforme

#### Sujet 1 :

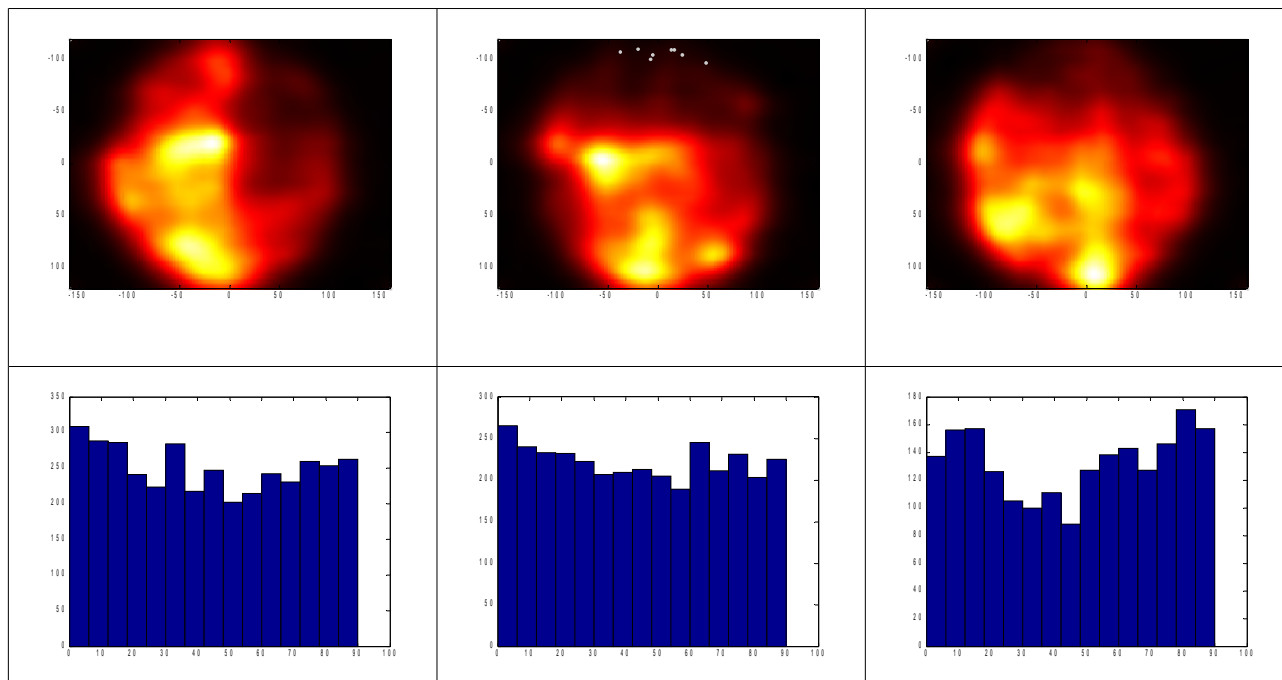


#### Sujet 2 :

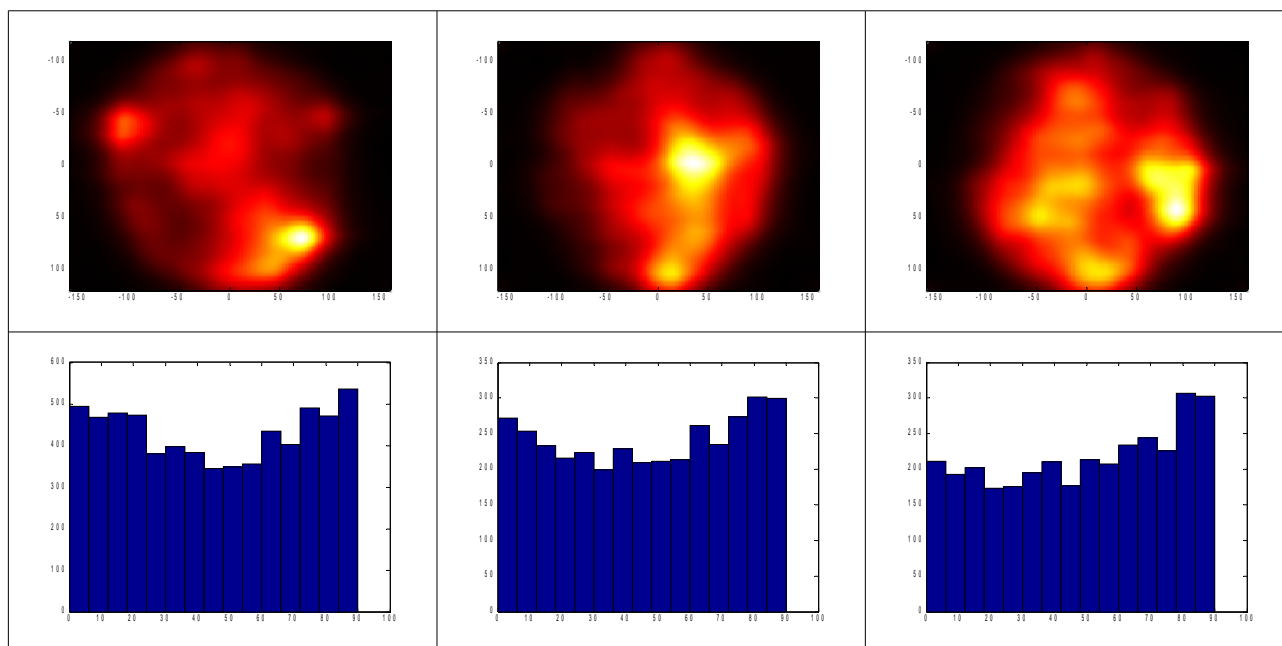




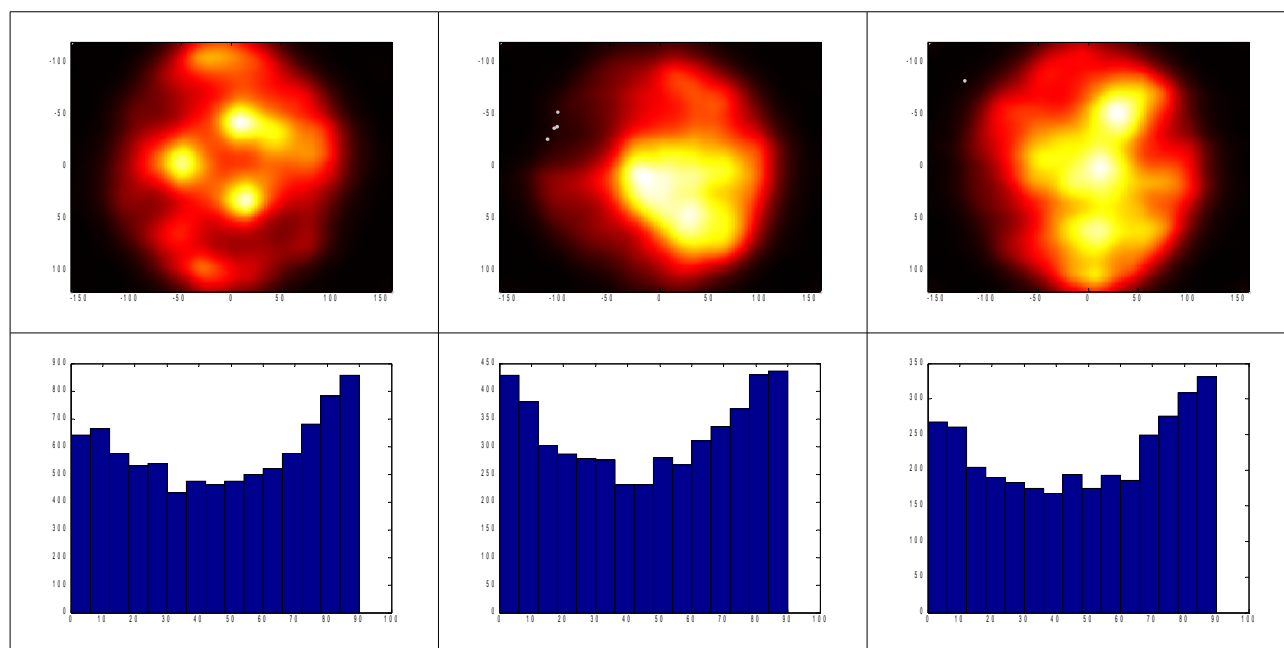
## Sujet 5 :



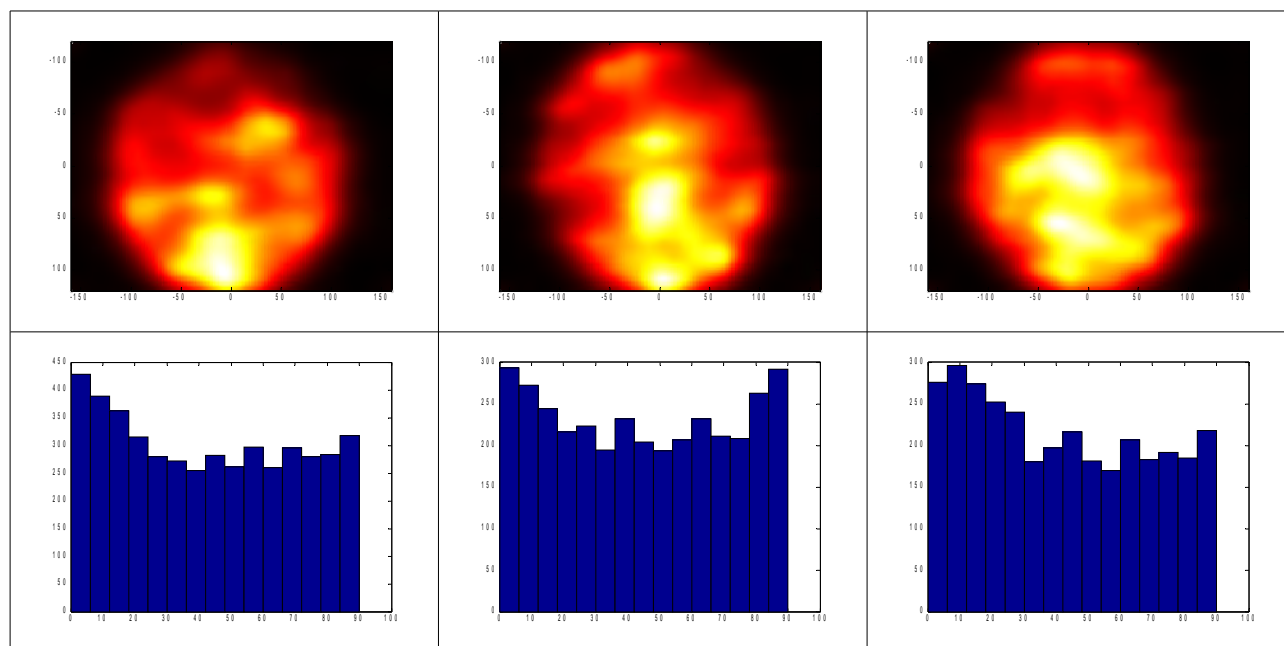
## Sujet 6 :



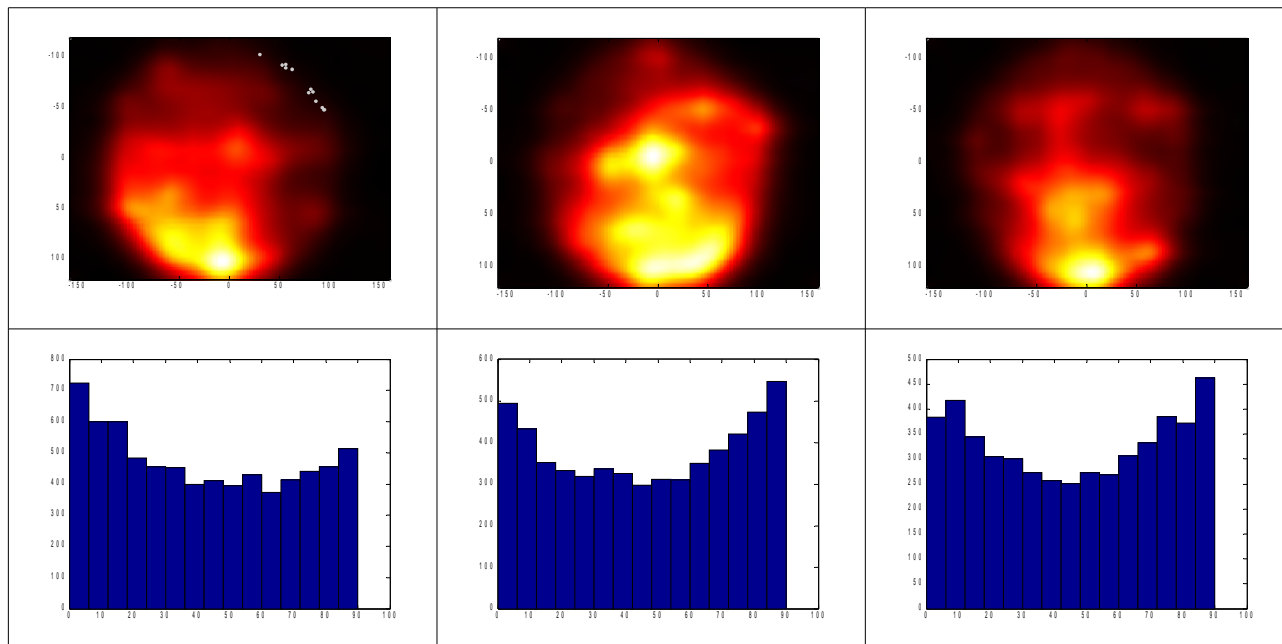
## Sujet 7 :



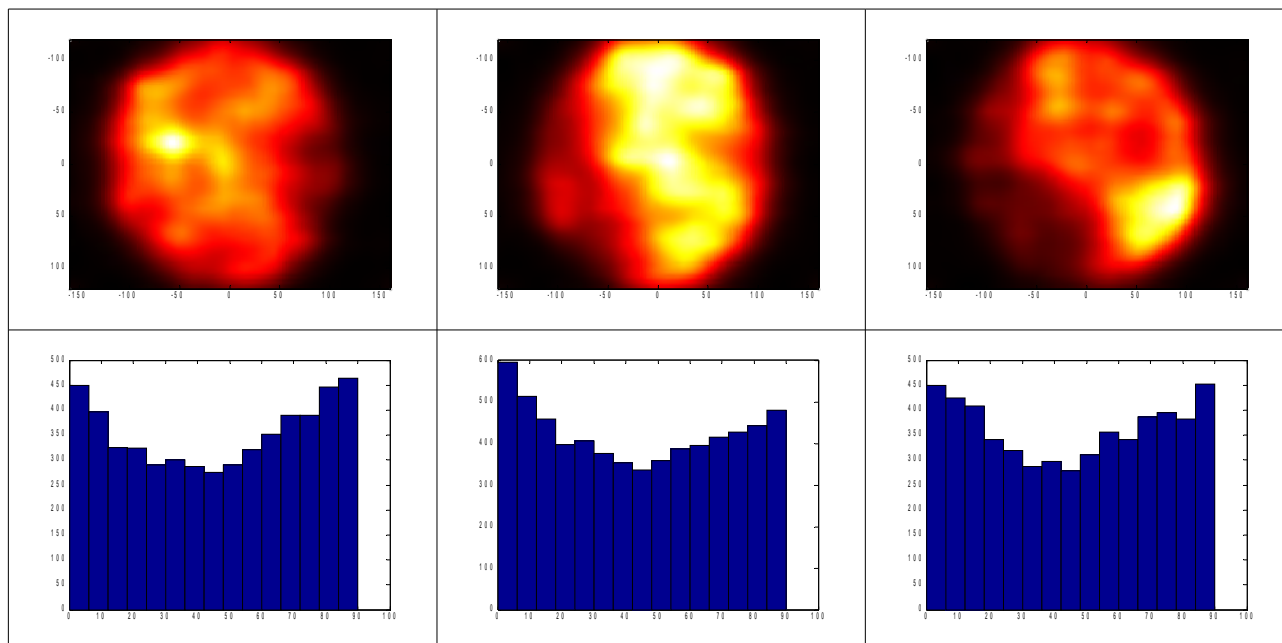
## Sujet 8 :



## Sujet 14 :

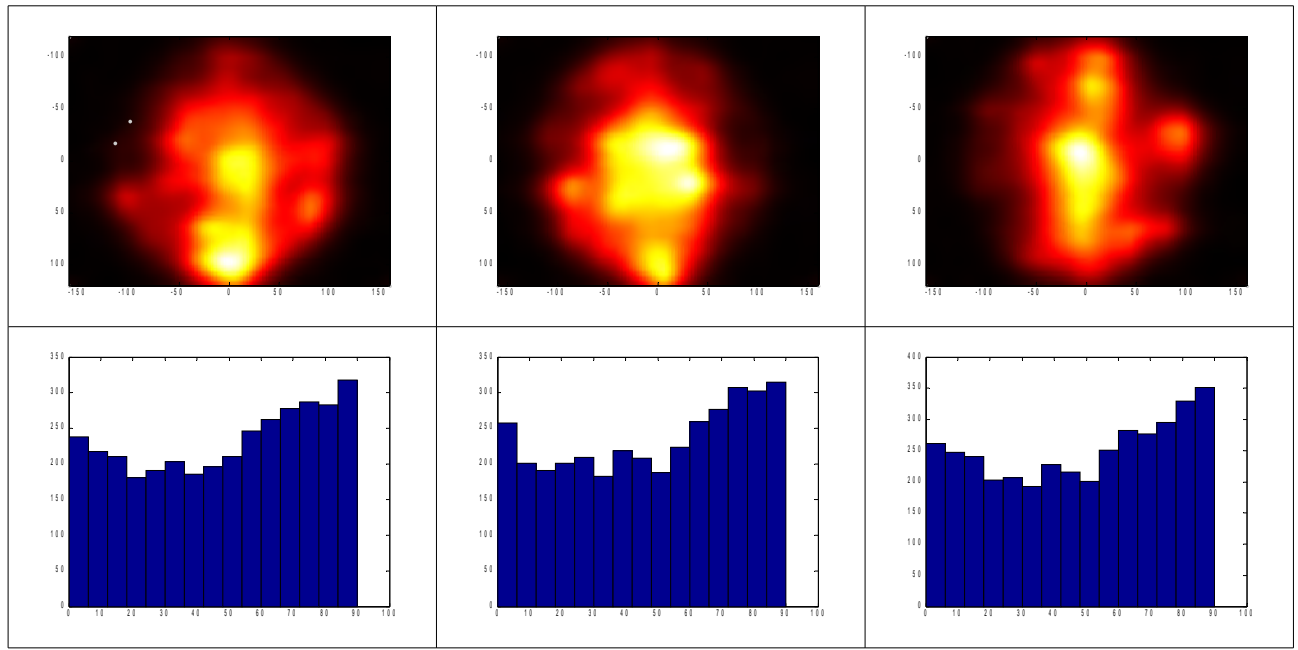


## Sujet 16 :

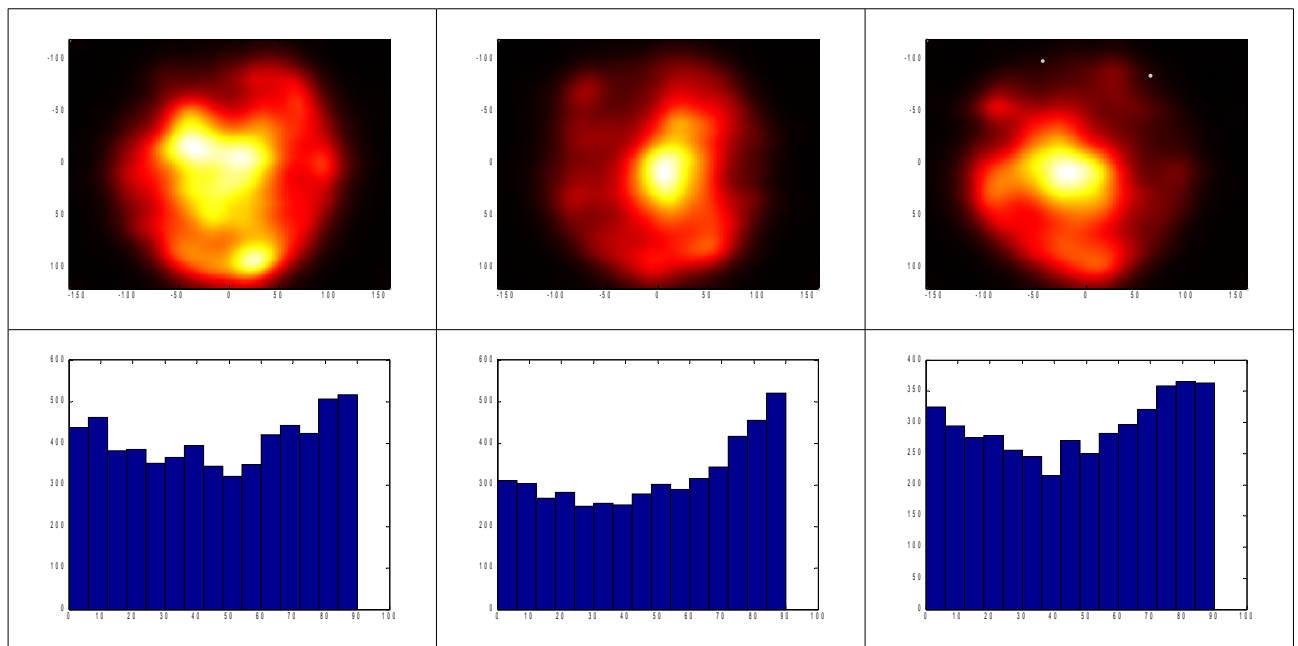


## Groupe LOG

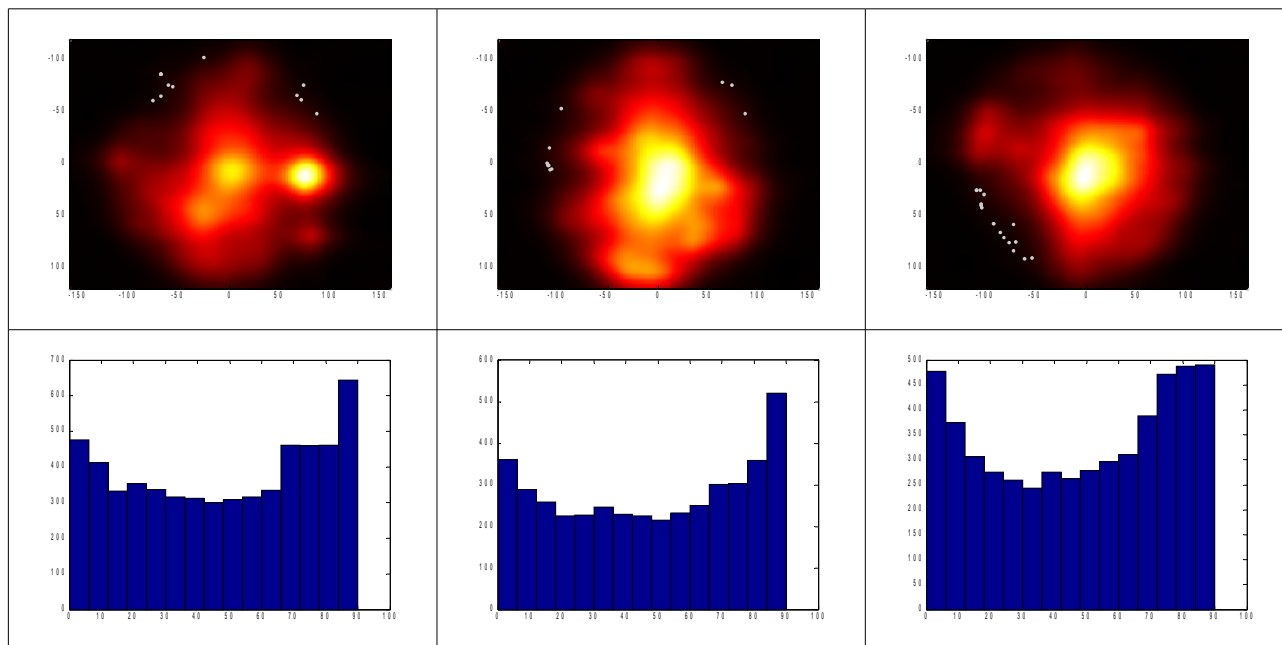
### Sujet 9 :



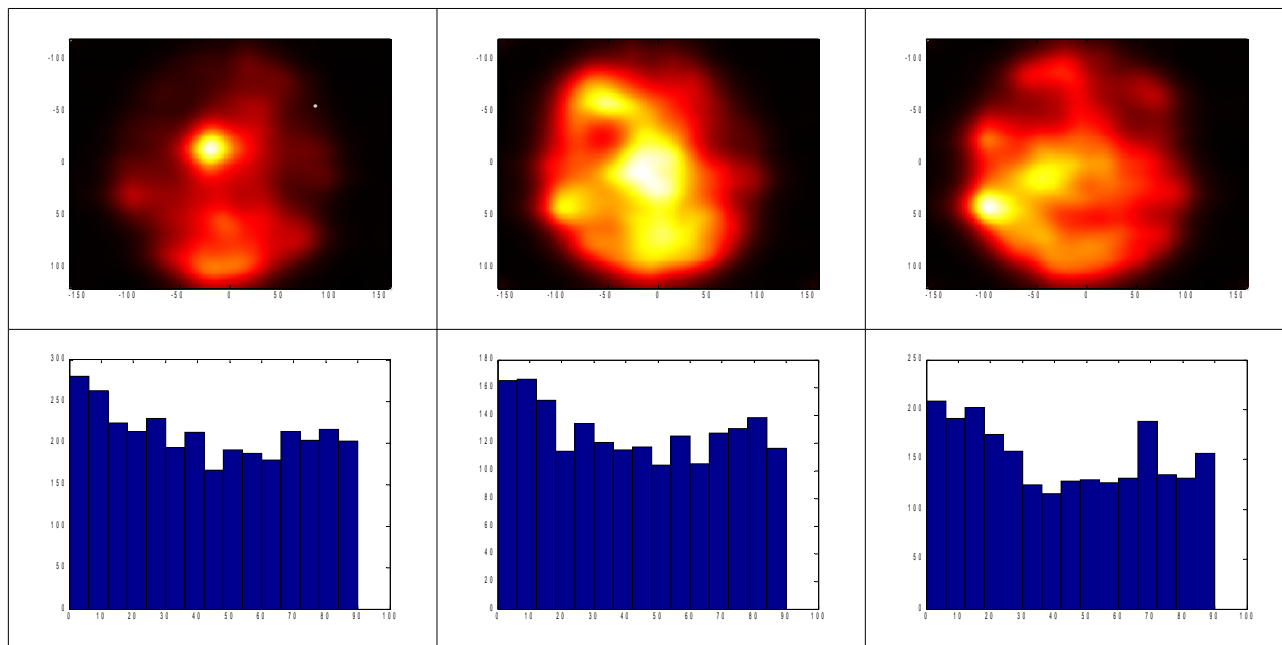
### Sujet 10 :



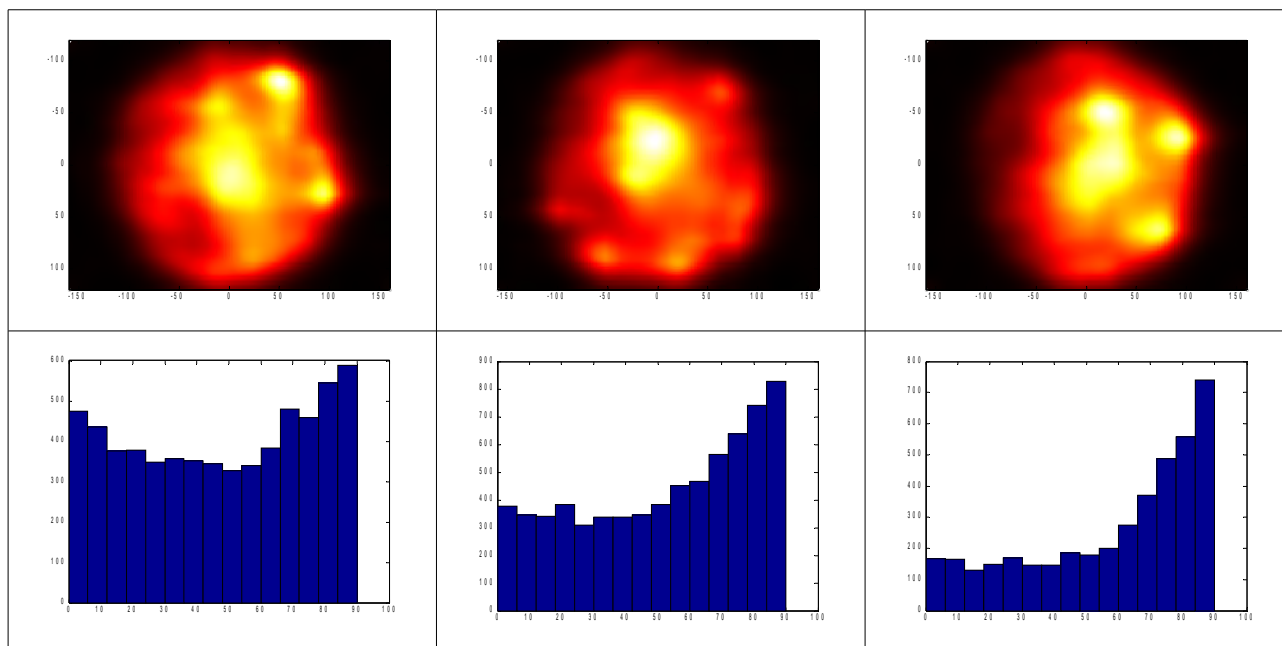
## Sujet 11 :



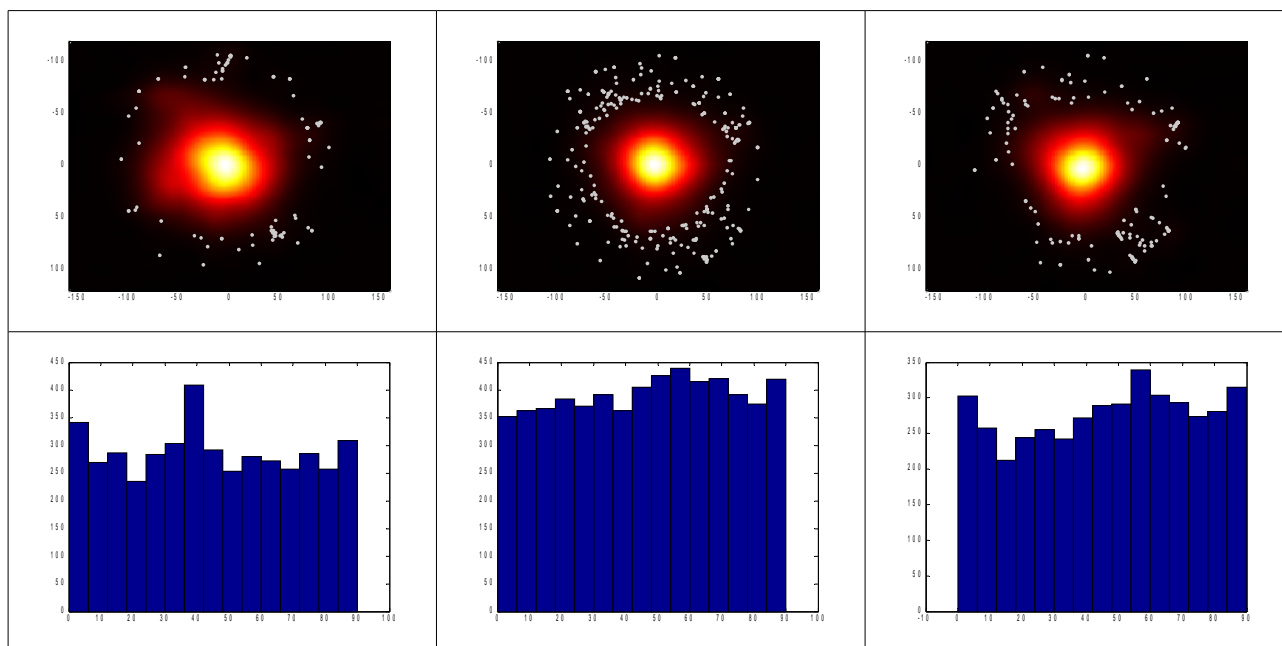
## Sujet 12 :



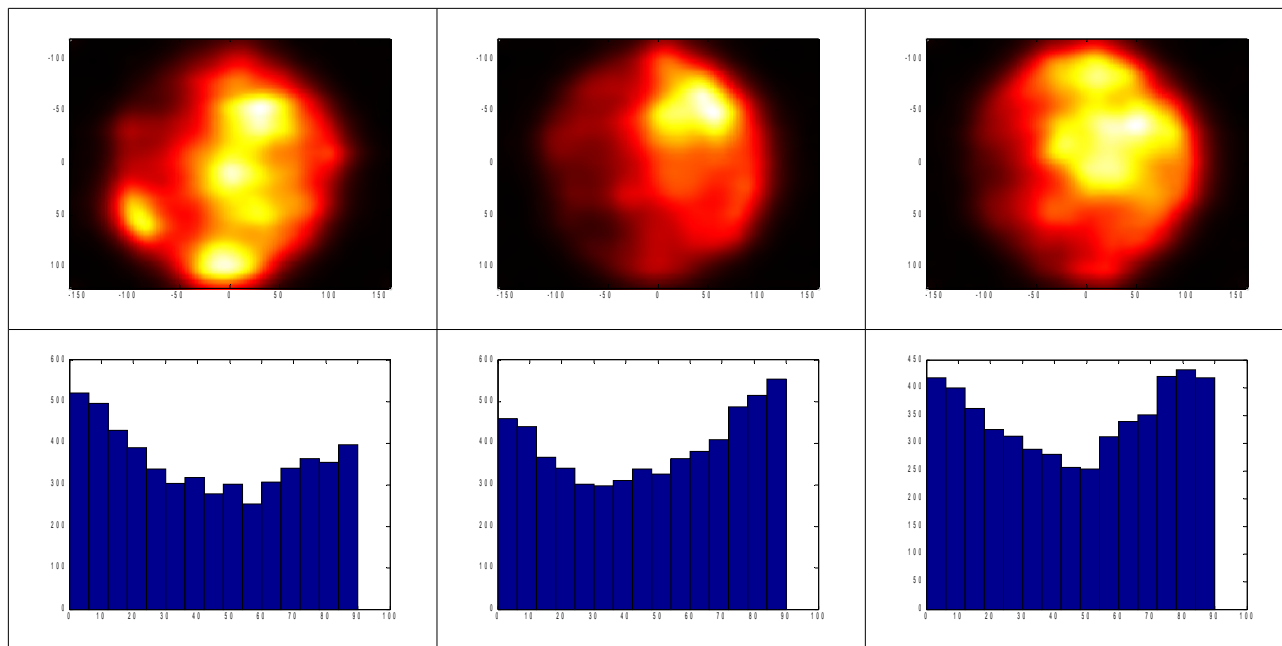
## Sujet 13 :



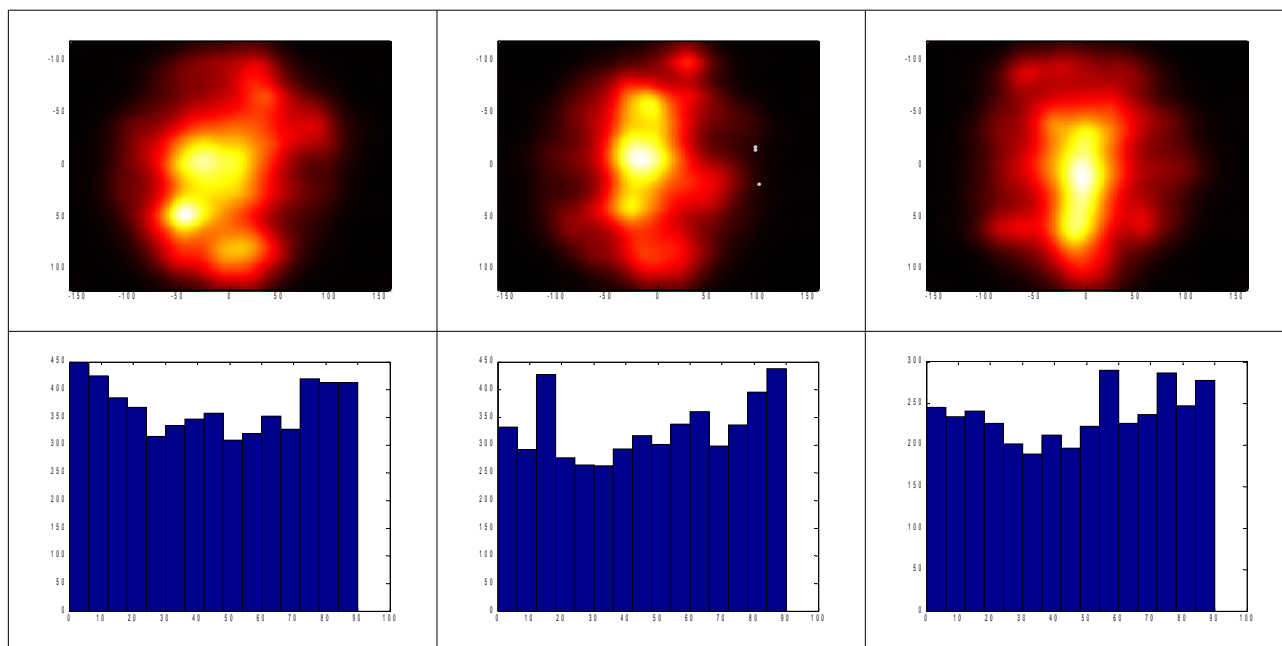
## Sujet 15 :



## Sujet 17 :



## Sujet 18 :



## Annexe du chapitre 5

### Réponse aux questionnaire de la 2ème expérience du chapitre 5

#### Sujet 1

Prénom : Bertrand

Age : 21

Sexe: M

Pratique musicale (instrument/durée) : non

Connaissance antérieure de la cour (oui/non) : non

Pratique des jeux vidéo (heures par semaine) : plus maintenant, mais il y a quelques années: >10h/semaine

#### Questionnaire expérience de mobilité.

1. L'apprentissage était-il trop long ou trop court ?

La première fois, oui, car on ne sait pas trop quoi regarder, on n'a aucune intuition sur le fonctionnement du système, donc ce serait peut-être bien de le faire une fois rapidement, et une seconde fois où on peut s'arrêter à volonté.

2. Y a-t-il des choses que tu aurais aimé faire avec le dispositif et que tu n'as pas pu faire ? Pourquoi ?

Faire fonctionner le système sans avoir les yeux bandés, mais je n'ai pas demandé à la fin de la dernière séance.

3. Comment décrirais-tu l'environnement dans lequel tu as évolué ?

4. Quels repères / stratégies as-tu utilisés pour te déplacer ?

Pour rester dans la même direction, j'essayais de retenir le son qui correspondait. En tournant la tête, j'essayais de déterminer ce qui me semblait le plus correspondre à ce que j'avais « enregistré » comme étant la direction.

Pour repérer les virages : je n'arrivais pas le premier, même en essayant de garder la tête tournée pour reconnaître la direction. Pour le second, il se trouvait à l'ombre et le changement dans le son, qui devenait plus sourd, m'indiquait le moment où je devais tourner.

5. A ton avis, comment fonctionne la conversion de l'image en son ?

J'en ai aucune idée. Je sais que ça fonctionne à partir de la luminosité. J'imagine que ça doit aussi travailler quelque part avec ce qui se passe avec la représentation dans le domaine de Fourier de l'image pour déterminer les fréquences du son. J'ai remarqué que les zones lumineuses étaient plus aiguës et plus forte (en regardant le ciel par exemple), et les zones d'ombres plus sourdes, et un peu plus basses en volume (la variation de volume est elle réelle ou simplement due à la perception des basses et des aiguës, je ne sais pas).

6. As tu ressenti le besoin de te protéger ?

Non

7. Est-ce que c'était difficile de se concentrer sur les sons ?

Pas vraiment difficile de se concentrer, mais je n'ai pas l'oreille musicale donc c'est surtout difficile à



interpréter les variations, et surtout reconnaître un ton quand on ne peut pas le comparer directement avec ce qu'on cherche.

8. Est-ce que tu aurais aimé faire plus de pauses ou que l'on coupe le son plus souvent?

Le son est moins gênant que le bandeau sur les yeux.

9. As tu eu l'impression qu'il y avait un décalage entre ce que tu voulais faire et ce que tu faisais réellement ?

Oui, beaucoup de mal à marcher droit, apparemment (je me suis retrouvé dans l'herbe après quelque pas seulement)

C'est aussi difficile de se réorienter une fois qu'on s'est trompé.

### **Commentaires ?**

C'était une expérience intéressante mais frustrante étant donné que je n'ai pas réussi à cerner le fonctionnement du système et à réaliser le parcours sans aide extérieure (mais y en a t il qui y arrive après seulement 4 séances ?)

Je suis assez étonné que ça ne fonctionne pas en détection d'objet, mais c'est vrai que ce serait difficile à représenter de façon sonore, étant donné qu'avec des écouteurs, on est limité dans l'interprétation spatiale de la source sonore (si je me trompe pas, avec un simple stéréo, on peut placer la source que dans une dimension, gauche ou droite, mais ce n'est pas possible de dire haut ou bas, devant/derrière, à moins d'utiliser un modèle compliqué qui simule la réponse de l'oreille externe).

Je me demande si en connaissant le fonctionnement du système on arriverait mieux.

D'ailleurs l'explication du processus est-elle impossible à un aveugle ?

## Sujet 2

Prénom : Thierry

Age : 43

Sexe: M

Pratique musicale (instrument/durée) :

Connaissance antérieure du parking (oui/non) :

Pratique des jeux vidéo (heures par semaine) :

### **Questionnaire expérience de mobilité.**

1. L'apprentissage était-il trop long ou trop court ?

Plutôt court.

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ? Pourquoi ?

Heu non

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé?

J'ai essayé de repérer un son spécifique.. et j'ai tenté de le suivre

5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

Aucuns.. j'ai appliqué les consignes .. qui étaient de ne se repérer qu'au travers les sons de la caméra

6. A votre avis, comment fonctionne la conversion de l'image en son ?

Je n'en sais rien du tout.. peut etre des sons différents en fonction du type d'objet que la camera 'voit' et de leur distance

7. Avez-vous ressenti le besoin de vous protéger ?

Non, la caméra faisait pare-chocs.. heureusement !!!

8. Etait-il difficile de se concentrer sur les sons ?

Pour etre franc.. oui

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Bien sur sinon je n'aurais pas fais de fautes.. rire !!!

### **Commentaires ?**

## Sujet 3

**Prénom : Thierry**

**Age : 43**

**Sexe: M**

Pratique musicale (instrument/durée) :

Connaissance antérieure du parking (oui/non) :

Pratique des jeux vidéo (heures par semaine) :

### Questionnaire expérience de mobilité.

1. L'apprentissage était-il trop long ou trop court ?

Plutôt court.

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ?  
Pourquoi ?

Heu non

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé?

J'ai essayé de repérer un son spécifique.. et j'ai tenté de le suivre

5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

Aucuns.. j'ai appliqué les consignes .. qui étaient de ne se repérer qu'au travers les sons de la caméra

6. A votre avis, comment fonctionne la conversion de l'image en son ?

Je n'en sais rien du tout.. peut etre des sons différents en fonction du type d'objet que la camera 'voit' et de leur distance

7. Avez-vous ressenti le besoin de vous protéger ?

Non, la caméra faisait pare-chocs.. heureusement !!!

8. Etait-il difficile de se concentrer sur les sons ?

Pour etre franc.. oui

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Bien sur sinon je n'aurais pas fais de fautes.. rire !!!

**Commentaires ?**

## Sujet 4

Voilà mes réponses

52 masculin

Pratique musicale piano, ( composition musique 30 ans classique <http://dofala.free.fr> des morceau mp3 )  
parking ????

Jeux video NON

1 : ni l'un ni l'autre ...

J'ai essayé plusieurs méthodes ...

2 : J'aurais aimé parcourir un cercle , avec des obstacles représentatif de l'univers sonore pour essayer de "capter"

3 : J'ai essayer de trouver des repères sonores dans les basses, les aigues et je n'ai eu qu'une petite idée pour un "obstacle" à l'arrivée

4 : J'ai obter pour un genre de balayage (scanning) avec la tete horizontalement

5 : J'ai essayé une fois sans balayage.

Et pour moi le balayage (meme inutile) me permettait de croire que j'accrocherais un "truc".  
et d'entendre une relation entre le mur de l'immeuble de gauche et le son (au départ)

6 : Vu le peu de complexité du matériel,  
je pense qu'il s'agit d'un onde modulé dans la bande sonore de 50hz à 17kh (par exemple)  
qui se réfléchit (il s'agirait de l'effet doppler)

7 : Aucun besoin de me protéger.

## Sujet 5

Prénom : Olivier Age : 25 ans Sexe: homme  
 Pratique musicale (instrument/durée) : piano 9 ans  
 Connaissance antérieure du parking (oui/non) : non  
 Pratique des jeux vidéo (heures par semaine) : non

### Questionnaire expérience de mobilité.

1. L'apprentissage était-il trop long ou trop court ?

Suffisamment long

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ? Pourquoi ?

J'aurais aimé l'apprivoiser librement en ma baladant n'importe où.

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué ?

Un brouhaha auditif, une pelote de sons, mais dont se dégagent des tendances et des variations parfois reconnaissables... On sent une évolution fréquentielle qui traduit la mobilité.

A l'ouïe s'ajoute le toucher (les pieds) auquel on prête plus attention qu'à l'habitude.

4. Comment vous êtes-vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisés ?

Au début je tournais la tête dans tous les sens pour « regarder » autour de moi mais ensuite j'ai plutôt gardé la même position (le regard légèrement en dessous de l'horizon il me semble) pour pouvoir reconnaître l'endroit. A la fin j'arrivais à reconnaître certains passages assez bien. Par contre si j'avais le malheur de me planter je me retrouvais vite perdu dans un environnement auditif inconnu et ininterprétable.

5. Quels autres repères/stratégies avez-vous utilisés pour réaliser le parcours ?

Les pieds ont parfois été utiles pour reconnaître certains passages. Et dans une moindre mesure les passages ombre-soleil : avec la chaleur qu'il faisait ça se ressent !

6. A votre avis, comment fonctionne la conversion de l'image en son ?

Il y a un effet stéréo lorsqu'on tourne la tête (si l'image bouge à droite, le son également). Les variations de fréquence m'ont donné parfois l'impression de rentrer dans une grotte notamment il me semble lorsque je passais à l'ombre.

Donc il me semble que chaque canal RVB est traduit en fréquence (d'où les basses dans l'ombre ?) et l'effet stéréo est dû à une détection de mouvement général dans l'image.

7. Avez-vous ressenti le besoin de vous protéger ?

Non

8. Était-il difficile de se concentrer sur les sons ?

Au début oui.

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Bof, surtout une impression de décalage les premières fois entre le parcours guidé et le parcours seul !

### Commentaires ?

Peut-être prévoir une séance pour laisser le candidat « jouer » avec le système ?

Le son brut peut-il être transformé de manière à mieux traduire l'espace ?

Gommer ce tissage fréquentiel un peu perturbant. Mais l'avantage d'un tel son brut est qu'avec l'expérience il peut sûrement être interprété dans n'importe quel contexte...

Améliorer l'ergonomie du dispositif mais bon ça n'est pas encore la priorité je pense...

Bon courage pour la suite !

## Sujet 6

Prénom : Sophie Age : 36 Sexe: F

Pratique musicale (instrument/durée) : non

Connaissance antérieure du parking (oui/non) : oui

Pratique des jeux vidéo (heures par semaine) : non

### Questionnaire expérience de mobilité.

1. L'apprentissage était-il trop long ou trop court ?

*Trop court. 2 jours de plus par exemple ou des séances un peu plus longues*

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ?  
Pourquoi ?

*Refaire le parcours sans le masque avec la caméra pour entendre et comprendre le fonctionnement.*

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

*Harmonieux, métallique et stressant*

4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé?

*Direction droite les bruits étaient plus métalliques, direction gauche plus sourds, tout droit comme un rythme ou une aspiration ...*

5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

*Le changement de sol, les différences de niveau, la chaleur du soleil.*

6. A votre avis, comment fonctionne la conversion de l'image en son ?

*J'aimerais bien le savoir !*

7. Avez-vous ressenti le besoin de vous protéger ?

*Tres souvent*

8. Etait-il difficile de se concentrer sur les sons ?

*Oui car il fallait aussi pouvoir entendre Nicolas nous rediriger et non car on avait que ça*

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

*Aucune idée mais une fois j'ai dû me diriger vers où je ne pensais pas devoir aller ...*

### **Commentaires ?**

***Nous avons bcp de chance d'avoir de bons yeux ! Nous utilisons d'autres sens sans s'en apercevoir. J'espère que ce projet aboutira et servira bcp de personnes.***

## Sujet 7

Prénom : Franck

Age : 22

Sexe: M

Pratique musicale (instrument/durée) : non

Connaissance antérieure du parking (oui/non) : non

Pratique des jeux vidéo (heures par semaine) : 2

### Questionnaire expérience de mobilité.

1. L'apprentissage était-il trop long ou trop court ?

Un peu court, trop en tout cas pour essayer toutes les possibilités.

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ?  
Pourquoi ?

J'aurais bien aimé savoir comment il marche et réessayer ensuite...

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

Seule l'ouïe était disponible, on ne pouvait s'aider de rien d'autre pour avancer, ce qui est assez perturbant puisqu'on ne se rend plus du tout compte de ce qu'on fait.

4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé?

J'essayais de trouver au sol la limite entre deux surfaces (en l'occurrence l'herbe et le goudron), quand c'était possible. Les 2 sons n'étant pas identiques, il faut voir quand ils changent, et rester le plus proche possible de la ligne de changement.

Les problèmes survenaient quand il fallait tourner ou que la limite n'était plus continue...

5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

J'ai essayé la première fois de compter le nombre de pas mais j'ai rapidement abandonné : trop dur et pas dans l'esprit de l'expérience il me semble. Sinon à part le son il n'y avait pas vraiment d'autres repères...

6. A votre avis, comment fonctionne la conversion de l'image en son ?

Je dirais qu'il y a comme paramètres principaux la luminosité (le bruit du soleil me semblait caractéristique) et la couleur (différence entre herbe et goudron). Peut-être la distance, ou la taille qu'occupe un objet dans le champ de vision. Après je ne sais pas du tout qui donnait quel son...

7. Avez-vous ressenti le besoin de vous protéger ?

Un peu, ça n'est jamais évident de marcher les yeux fermés.

8. Etait-il difficile de se concentrer sur les sons ?

Pas particulièrement, vu que c'est la seule chose que l'on peut faire. Il aurait par contre fallu plus de temps pour bien appréhender les nuances.

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Parfois oui, je pensais reconnaître quelque chose (banc, poubelle...) alors que visiblement ça n'était pas ça du tout. Sinon sur les derniers tests je pense que globalement j'arrivais à suivre le parcours.

### **Commentaires ?**

## Sujet 8

Prénom : Nicolas

Age : 22

Sexe: M

Pratique musicale (instrument/durée) : Aucun

Connaissance antérieure du parking (oui/non) : Non

Pratique des jeux vidéo (heures par semaine) : 30h

### Questionnaire expérience de mobilité.

1. L'apprentissage était-il trop long ou trop court ?

La durée d'apprentissage était légèrement trop longue (1 séance) étant donné que le lieu, et la disposition des éléments, étaient les mêmes.

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ? Pourquoi ?

J'aurais aimé pouvoir tester le son en sachant ce que je voyais afin d'affiner la visualisation. Par exemple, se mettre devant un arbre et écouter la différence. Pour une utilisation courante, il est intéressant de pouvoir faire la différence entre un arbre et une personne. (par exemple), et c'est quelque chose que je ne pouvais pas faire.

J'aurais aussi voulu tester la résolution qui fait que le son change. Sur le chemin, on pouvait faire la différence entre celui-ci et l'herbe, mais ne sachant pas la résolution, il est difficile d'évaluer la véritable distance avec l'angle de la tête (dans le cas d'une différence de niveau, cela peut poser problème). Une image noir et blanc qu'on passe devant la caméra, yeux ouverts, pourrait permettre d'entendre comment se fait un tel passage. Ou, si on veut le faire avec les yeux fermés, un objet en relief dont on peut sentir les 2 niveaux qui sont chacun d'une couleur bien différente.

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

L'environnement était très varié et il me semblait changer de complètement de lieu en regardant certaines choses.

Il était fait de faible variation, même lorsqu'on regardait un point fixe.

Malgré sa diversité, c'est un environnement difficile à comprendre parfaitement, à la fois à cause des faibles qualités audio de notre ouïe (qui rate facilement un très léger changement de fréquence) et d'autre part à cause de la difficulté de visualisation de cet environnement.

4. Comment vous êtes-vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisés?

Lors des premiers parcours, j'ai tenté de regarder face à moi, comme si je marchais normalement à vue. Bien qu'il y avait des différences, dans les sons, il était impossible d'accrocher un repère intéressant à suivre, "l'image" étant trop "floue" pour distinguer à longue distance.

Puis j'ai tenté la distance intermédiaire, la tête vers le bas, mais pas trop. Il était alors possible de repérer les textures unies sans trop de difficulté.

Au fur et à mesure, je me suis rendu compte qu'il valait mieux regarder très près de soi (1-2 mètres devant ses pieds) pour pouvoir véritablement savoir ce qu'il y avait devant soi.

Une fois cela acquis, j'ai distingué l'herbe du chemin afin de suivre le dernier en balayant de la tête ce qui se situait devant moins sur un angle d'environ 45°

De temps à autre, je levais la tête afin de repérer les arbres (ou ce que je pensais être des arbres) afin d'avoir une idée de l'avancement sur le chemin.

De plus, certains éléments du parcours avaient des bruits caractéristiques, comme la fin par exemple. Il y avait



là-bas un bruit qu'il n'y avait nul par ailleurs sur le chemin et plus on s'en approchait, plus la différence était sensible.

5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

N'étant privé que de la vue, je me suis permis d'utiliser le touché afin d'avoir des informations sur le sol. A certains endroits, le chemin présente des irrégularités dues au fait qu'il soit fait de dalle. En marchant dessus, il était possible d'avoir un repère d'orthogonalité avec le chemin permettant de se remettre droit le cas échéant.

En longeant d'assez près l'herbe, il était possible d'y aller uniquement en faisant attention à la petite différence de niveau entre le chemin et l'herbe. Elle est pas énorme, mais elle se remarque facilement si on y prête attention.

Bien entendu, après un certain nombre de passages, la mémoire permet de se souvenir de point important comme le fait qu'il faut tourner à droite une fois et un peu plus loin sur la gauche. Je n'ai pas essayé d'aller jusqu'à compter le nombre de pas, mais au bout d'un moment, il était possible d'avoir l'intuition qu'il fallait tourner où que j'étais allé trop loin.

6. A votre avis, comment fonctionne la conversion de l'image en son ?

Je pense qu'il y a une relation directe avec la luminance de l'image et la fréquence du son. Regarder le ciel (forte luminance) rendait le son cristallin (plutôt aigue) alors que regarder de près un mur sombre rendait le son caverneux (grave).

On remarquait une différence du même ordre entre l'herbe et le chemin.

Je dirais qu'une moyenne quelconque de la luminance donne la fréquence principale du son.

Puis, les variations sonores proviendraient des variations à cette moyenne sur un principe de DCT en sélectionnant certains coefficients pour donner des informations horizontal ou vertical. Les coefficients verticaux augmentant la fréquence maximal et les coefficients horizontaux diminuant la fréquence minimal (ou inversement).

7. Avez-vous ressenti le besoin de vous protéger ?

Au tout début, oui ! Mais très rapidement, le besoin de se protéger disparaît. D'une part parce qu'il gêne (on me facilement les mains dans le champ de vision de la caméra) et d'autre part parce qu'on prend confiance dans ce qu'on ressent.

8. Etait-il difficile de se concentrer sur les sons ?

La première fois, c'était difficile. Les sons sont étranges et on arrive pas à les utiliser efficacement. Ils semblent alors inutiles et on se concentre alors peu dessus.

Mais une fois qu'on a trouvé comment interpréter ces sons, la concentration est facile, même avec du bruit autour.

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

J'ai l'impression que je ne marchais pas droit. Même quand je suivais l'herbe, j'avais l'impression de m'en éloigner en entendant le son changé sans avoir bougé la tête.

**Commentaires ?**

Le système est, je pense, efficace dans un environnement connu, fait et refait mais je pense qu'il aurait été intéressant de faire au moins une séance à la fin en changeant le lieu.

La connaissance précise du chemin à parcourir biaise légèrement le résultat final car on s'attend à devoir faire une certaine action à un certain moment.

## Sujet 9

Prénom : Christophe

Age : 45

Sexe:

masculin

Pratique musicale (instrument/durée) : aucune

Connaissance antérieure du parking (oui/non) : non

Pratique des jeux vidéo (heures par semaine) : 3 heures

### **Questionnaire expérience de mobilité.**

1. L'apprentissage était-il trop long ou trop court ?

La privation de l'un des sens, désoriente le sujet. Plus cette phase est courte, et meilleur est l'apprentissage, puisque que le sujet est presque demandeur de cette compensation de perte de sens. Dans mon cas, la phase d'apprentissage m'a parue correcte, compte tenu de la mise en place de l'appareillage, le repérage du parcours en aveugle.

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ? Pourquoi ?

Après une période d'apprentissage peut être plus longue, j'aurais aimé passer à proximité d'une zone perturbée (niveau de bruit ambiant urbain par exemple) afin de « voir » s'il était possible de maintenir sa concentration.

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

Cette expérience m'a fait penser à certains épisodes de « cosmos 1999 », c'est-à-dire, perte de repère et obligation d'une nouvelle concentration sur les éléments environnants : le son. Il n'y avait pas d'angoisse ni d'appréhension par rapport à l'équilibre et le mode de déplacement la marche.

4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé?

J'ai essayé de trouver une relation auditive, percevoir les variations du sons afin d'en établir des codes directionnels.

5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

Les sensations du pied avec le sol m'ont parfois donné des indications de directions (bordures d'herbes, trous, etc...).

6. A votre avis, comment fonctionne la conversion de l'image en son ?

Sur la base d'une graduations des contrastes lumières et retranscriptions en sons

7. Avez-vous ressenti le besoin de vous protéger ?

Aucunement, d'ailleurs, j'ai gentiment flirté avec les poubelles

8. Etait-il difficile de se concentrer sur les sons ?

Une fois l'appareillage posé, la mise en condition est instantanée

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Oui, légère et probablement générée par une inconsciente hésitation

**Commentaires ?**

Expérience très intéressante, mais un appareillage plus compact et enveloppant (casque intégral ultra léger) rendrait peut être l'expérience plus réaliste encore (ne pas ressentir le moyen de privation du sens).

## Sujet 10

Prénom : Jonathan

Age : 24

Sexe: Masculin

Pratique musicale (instrument/durée) : Saxophone (3 ans)

Connaissance antérieure du parking (oui/non) : non

Pratique des jeux vidéo (heures par semaine) : oui (10+h/semaine)

### **Questionnaire expérience de mobilité.**

1. L'apprentissage était-il trop long ou trop court ?

Peut être un petit peu trop court mais dans l'ensemble la durée était bonne

1)2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ? Pourquoi ?

Aller autre part que dans le parking... Pour voir comment ça aurait réagi.

2)3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

? Un son un peu "bruit de phase" plus ou moins rapide suivant ou l'on "écoute". Difficile d'entendre autre chose que ce bruit par contre, si on le met assez fort pour bien saisir les variations.

3)4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé?

Essayer de repérer le "son" du chemin. Essayer de repérer les "sons" de ce qui n'est pas le chemin au cas où...

En faisant aussi attention à l'ombre : le chemin dans l'ombre n'a pas le même son que le chemin au soleil. On sent quand on est au soleil ou dans l'ombre donc une fois que l'on y est ça ne pose pas de problème mais quand on est au soleil et que l'on regarde dans l'ombre ou l'inverse cela peut faire douter que l'on est sur le chemin.

4)5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

Essayer de mémoriser les grandes lignes du trajet (l'ordre des directions). Et s'arrêter de temps en temps pour "écouter" l'endroit où l'on est (autour).

5)6. A votre avis, comment fonctionne la conversion de l'image en son ?

La fréquence du bruit suivant la luminosité de l'image et sûrement beaucoup de choses en plus...

6)7. Avez-vous ressenti le besoin de vous protéger ?

Non.

7)8. Était-il difficile de se concentrer sur les sons ?

Non. Le son n'est pas trop agressif.

8)9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Non, pas en marchant en tout cas. Cela peut arriver si on bouge trop vite il me semble.

### **Commentaires ?**

J'aurai bien aimé l'utiliser autre part que sur le parking^^.

## Sujet 11

Prénom : *Matthias*

Age : *21ans*

Sexe: *masculin*

Pratique musicale (instrument/durée) : *clarinette/15ans*

Connaissance antérieure du parking (oui/non) : *non*

Pratique des jeux vidéo (heures par semaine) : *non*

### **Questionnaire expérience de mobilité.**

1. L'apprentissage était-il trop long ou trop court ?

*L'apprentissage était un peu court, pour moi il aurait plus efficace en intensif sur 2 jours*

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ?  
Pourquoi ?

*Me promener dans un endroit familier et dans un lieu avec des personnes*

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

*Un peu irréel, coupé du monde extérieur, déambulant sans utiliser les expériences sensorielles habituelles*

4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé?

*J'ai utilisé plusieurs techniques les première fois puis j'ai trouvé la plus appropriée pour moi : en regardant le sol devant moi puis la tête plus relevée sur la fin du parcours.*

5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

*Balayage droite gauche entre autre, tête fixe en « regardant » en face.*

6. A votre avis, comment fonctionne la conversion de l'image en son ?

*Je n'en ai aucune idée*

7. Avez-vous ressenti le besoin de vous protéger ?

*Non*

8. Etait-il difficile de se concentrer sur les sons ?

*Oui car il était très nombreux, et par conséquent brouillés.*

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

*Non pas spécialement.*

### **Commentaires ?**

*Je vois assez mal ce système développé pour des malvoyants car il coupe du monde extérieur un peu plus que le handicap encore. Est-ce que le système marche de nuit du fait que le son provient de l'image d'une caméra ?*

## Sujet 12

Prénom : Chloe Age : 23 ans Sexe: F

Pratique musicale (instrument/durée) : Piano / 10 ans

Connaissance antérieure du parking (oui/non) : non

Pratique des jeux vidéo (heures par semaine) : 0

### **Questionnaire expérience de mobilité.**

1. L'apprentissage était-il trop long ou trop court ?

Trop court pour moi, je ne pense pas avoir réussi à me familiariser suffisamment avec le dispositif, repérer par exemple les différents sons et les utiliser pour pouvoir m'orienter.

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ?  
Pourquoi ?

Faire le parcours une fois avec vision + son, en fin d'expérience, pour voir par exemple quelle vue produit quel son (mais simplement par curiosité)

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

Assez surprenant, inhabituel : comme les bruits « naturels » (passants, voiture, etc.) sont assez masqués par les sons, ça donne un environnement « fermé » dans le sens où on ne peut se repérer qu'à l'aide de ces sons particuliers et nouveaux

4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé?

En essayant de repérer des sons particuliers au moment où je devais dévier à gauche/droite (= quand je faisais une erreur/obstacle, comme une sortie de parcours par exemple), et pendant le chemin du retour (d'ailleurs, je n'avais pas le son pendant un des parcours retour, ce qui a changé ma perception du parcours)

5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

Même en essayant de me baser sur les sons au maximum, je pense aussi m'être aidée de la distance parcourue, du temps, par exemple me rappeler qu'après une 20aine de secondes je déviais à gauche ou à droite les fois précédentes

6. A votre avis, comment fonctionne la conversion de l'image en son ?

Très bonne question... désolée, mais je n'en ai aucune idée, vraiment trop abstrait pour moi !

7. Avez-vous ressenti le besoin de vous protéger ?

Envie de mettre les mains en avant, reculer la tête ou fermer les yeux, mais le premier jour uniquement

8. Etait-il difficile de se concentrer sur les sons ?

Le son étant réglable facilement, pas de problème de ce côté-là (mais parfois envie de « séparer » un son particulier du reste)

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Non, mais impression qu'au final je n'ai pas fait ce que j'avais l'impression de faire (par exemple, j'imaginais

qu'il y avait des déviations plus importantes avant de voir le parcours).

### **Commentaires ?**

Est-ce que les résultats sont bien moins bons sans dispositif ? C'est vraiment fort si on est capable d'utiliser un tel dispositif, qui est vraiment nouveau pour nous, en si peu de temps !

Est-ce qu'on est capable de généraliser ? Si, une fois le parcours x appris, on nous pose avec le dispositif sur un parcours y on est capable de se débrouiller ?

Pendant le parcours, je n'avais pas l'impression d'être dans le noir, c'est-à-dire comme si un son appelait une image, une couleur (non, je ne me drogue pas ☐). Est-ce que ça peut être le cas pour des personnes non voyantes ?

Détail (surement inutile) : je n'ai absolument aucun sens de l'orientation (en fait, c'est limite pathologique chez moi !) donc ça a peut-être joué sur l'apprentissage.



## Sujet 13

Prénom : Chloé Age : 24 Sexe: F

Pratique musicale (instrument/durée) : piano/ 6 ans mais il y a longtemps

Connaissance antérieure du parking (oui/non) : non

Pratique des jeux vidéo (heures par semaine) : oulala... des heures mais encore il y a longtemps

### **Questionnaire expérience de mobilité.**

- 1)1. L'apprentissage était-il trop long ou trop court ? juste ce qu'il faut
- 2)2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ?  
Pourquoi ? Tester avec un peu plus d'attention les effets de l'éclairage (luminosité) pour détecter les différences et les variations de la représentation d'un même objet avec plus de détails.
- 3)3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué? Cet environnement ressemble à une vision virtuelle de l'environnement qui nous entoure où chaque chose dans l'espace est représenté par un son particulier dans la bande de fréquence audible. Mais il est parfois difficile de distinguer les nuances du fait que ce système est peut-être un peu trop sensible à la luminosité de l'instant d'utilisation.
- 4)4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé(e)s?  
Recherche des bords, de surfaces lisses et des paquerettes ;-). Analyse des sons connus par apprentissage, facilement détectables et reconnaissables même si ils sont quelques fois difficilement discernables du bruit de fond. (distinction entre le chemin et l'herbe, détection du pylône de fin, etc...)
- 5)5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours? Le son extérieur (le tram qui passe !!) et toi qui nous remets dans le droit chemin !!
- 6)6. A votre avis, comment fonctionne la conversion de l'image en son ? chaque point de l'image est converti en son dans la bande de fréquence audible selon sa position et son intensité. Une notion de mouvement est intégrée au système.
- 7)7. Avez-vous ressenti le besoin de vous protéger ? OUI !!!! il y a des endroits qui font peur !!!! (zone d'ombre où l'on a l'impression que quelque chose nous tombe sur la tête) Puis le pylône de fin...
- 8)8. Etait-il difficile de se concentrer sur les sons ? Non, sauf lorsque l'on passe dans des zones trop bruitées (toujours du fait de la différence de luminosité, les sons changent quelques fois et sont difficilement distinguables)
- 9)9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ? Non
- 10)

**Commentaires ?** Expérience sympathique qui montre à quel point l'audition peut être exploitée et servir au repérage dans l'espace, même si selon moi un tel système a pour défaut principal l'obstruction des sons extérieurs indispensable à la vie sociale d'un non ou mal voyant utilisant ce système.

## Sujet 14

Prénom : Alice Age : Sexe:

Pratique musicale (instrument/durée) :

Connaissance antérieure du parking (oui/non) :

Pratique des jeux vidéo (heures par semaine) :

**Questionnaire expérience de mobilité.**

1. L'apprentissage était-il trop long ou trop court ?

non

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ?  
Pourquoi ?

Oui explorer objets simple + toucher

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

Pas compris

4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisé?  
Utiliser pour certains repères exemple batiments

5. Quels autres repères/stratégies avez-vous utilisé pour réaliser le parcours?

Repères auditif tram, calorimétrique ombre des arbres, proprioceptifs et vestibulaires.

6. A votre avis, comment fonctionne la conversion de l'image en son ?  
Sais pas

7. Avez-vous ressenti le besoin de vous protéger ?

Si j'allais trop vite

8. Etait-il difficile de se concentrer sur les sons ?

Non mais ce n'est pas utile

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Non

**Commentaires ?**

## Sujet 15

Prénom : Camille Age : 27ans Sexe: Féminin

Pratique musicale (instrument/durée) : percussions  
Connaissance antérieure du parking (oui/non) : ne s'applique pas  
Pratique des jeux vidéo (heures par semaine) : jamais

### Questionnaire expérience de mobilité.

10. L'apprentissage était-il trop long ou trop court ?

Trop court. Je ne crois pas avoir identifié comment fonctionnait le dispositif. Je n'ai pas reconnu le type de son émis en fonction du suivi correct du parcours.

11. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ?  
Pourquoi ?

J'aurais souhaité être avertie des obstacles par un son plus identifiable que ceux entendus. J'aurais voulu que le dispositif m'indique si le chemin allait plus droit ou plus à droite, à gauche.

12. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué ?

J'ai eu l'impression de marcher sur des dalles et d'être constamment au bord d'un mont de terre ou pelouse. Il y avait des arbres et les dalles formaient des courbes.

13. Comment vous êtes-vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisés ?  
J'ai essayé de me servir des sons. En fonction des sons entendus, je me dirigeais, je changeais de direction. Je me suis également servie des indications gauche-droite de Nicolas. En fonction de ça et des sons, je changeais ma position. Si les sons étaient plus forts, j'allais là où ils étaient moins assourdissants.

14. Quels autres repères/stratégies avez-vous utilisés pour réaliser le parcours ?  
J'ai essayé de me rappeler la direction de mes pieds, une sorte de repaire spatial en quelque sorte.  
Une fois j'ai essayé de compter mes pas mais en vain.

15. A votre avis, comment fonctionne la conversion de l'image en son ?

Quand il y a un obstacle, les sons sont plus forts. Ils ronronnent en quelque sorte quand vous allez dans la bonne direction.

16. Avez-vous ressenti le besoin de vous protéger ?

Oui. Des fois j'avais l'impression d'être face à un mur et de me le taper ou d'être au bord d'un vide.

17. Était-il difficile de se concentrer sur les sons ?

Oui, mais c'est aussi le fait que j'étais à la base déjà fatiguée, je pense.

18.

19. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Oui, tout à fait.

### **Commentaires ?**

**peut-être qu'en d'autres circonstances, j'aurais mieux compris...ceci restera une question non élucidée...**

ca plus une canne, ça devrait faire l'affaire pour s'orienter.

## Sujet 16

Prénom : Elodie

Age : 23

Sexe: F

Pratique musicale (instrument/durée) : flûte traversière (6 ans)

Connaissance antérieure du parking (oui/non) : non

Pratique des jeux vidéo (heures par semaine) : 0

**Questionnaire expérience de mobilité.**

1. L'apprentissage était-il trop long ou trop court ?

Plutôt trop court

2. Y a-t-il des choses que vous auriez aimé faire avec le dispositif et que vous n'avez pas pu faire ?  
Pourquoi ?

non

3. Pouvez-vous décrire l'environnement perceptif dans lequel vous avez évolué?

Déstabilisant, les sons étaient différents selon les endroits mais pas assez pour pouvoir se repérer.

4. Comment vous êtes vous servi de l'appareil ? Quels repères/stratégies avez-vous utilisés ?  
A chaque fois j'essayais de repérer un son particulier et l'endroit où je devais me situer par rapport à lui (soleil, bâtiment...)

5. Quels autres repères/stratégies avez-vous utilisés pour réaliser le parcours?

La mémoire... mais cela n'a pas bien marché !

6. A votre avis, comment fonctionne la conversion de l'image en son ?

J'imagine que les sons se différencient en fonction de la matière filmée par la caméra (végétaux, béton, ciel...)

7. Avez-vous ressenti le besoin de vous protéger ?

Non puisque je marchais doucement.

8. Etait-il difficile de se concentrer sur les sons ?

Non

9. Avez-vous eu l'impression qu'il y avait un décalage entre ce que vous vouliez faire et ce que vous faisiez réellement ?

Oui ! J'ai un très mauvais sens de l'orientation, parfois je pensais aller tout droit et en réalité je faisais des zigzag.

**Commentaires ?****Expérience intéressante. J'aurais aimé avoir plus d'informations sur le fonctionnement de l'appareil et ses applications possibles.****Bon courage !**